SUB-RIEMANNIAN AND SUB-LORENTZIAN GEOMETRY ON $\mathrm{SU}(1,1)$ AND ON ITS UNIVERSAL COVER

ERLEND GRONG AND ALEXANDER VASIL'EV

Abstract. We study sub-Riemannian and sub-Lorentzian geometry on the Lie group $\mathrm{SU}(1,1)$ and on its universal cover $\mathrm{CSU}(1,1)$. In the sub-Riemannian case we find the distance function and completely describe sub-Riemannian geodesics on both $\mathrm{SU}(1,1)$ and $\mathrm{CSU}(1,1)$, connecting two fixed points. In particular, we prove that there is a strong connection between the conjugate loci and the number of geodesics. In the sub-Lorentzian case, we describe the geodesics connecting two points on $\mathrm{CSU}(1,1)$, and compare them with Lorentzian ones. It turns out that the reachable sets for Lorentzian and sub-Lorentzian normal geodesics intersect but are not included one to the other. A description of the timelike future is obtained and compared in the Lorentzian and sub-Lorentzain cases.

Contents

2000 *Mathematics Subject Classification.* Primary: 53C17, 53B30, 22E30; Secondary: 53C50, 83C65.

Key words and phrases. SU(1,1), universal cover, optimal control, sub-Riemannian and sub-Lorentzian manifolds, Carnot-Carathéodory metric, geodesic.

This work was completed with the support of the grants of the Norwegian Research Council #177355/V30, of the European Science Foundation Research Networking Programme HCAA, and of the NordForsk Network 'Analysis and Applications'.

1. Introduction

Sub-Riemannian geometry is proved to play an important role in many applications, e.g., in mathematical physics and control theory. Sub-Riemannian geometry enjoys major differences from the Riemannian being a generalization of the latter at the same time, e.g., geodesics may be singular, the Hausdorff dimension is larger than the manifold topological dimension, the exponential map is never a local diffeomorphism. There exists a large amount of literature developing sub-Riemannian geometry. Typical general references are [19, 21, 22]. The sub-Lorentzian case is less studied and the first works in this directions appeared rather recenlty, see [8, 11, 12, 17].

In the development of sub-Riemannian geometry, one observes several examples, which are mainly nilpotent Lie groups, with either a left or right invariant distribution and metric. A sample representative is the Heisenberg group (see, e.g., [19]). Analysis of these groups in the sub-Riemannian setting, is already well studied. While these groups enjoy the advantage that explicit results are easier to obtain, their properties are sometimes too nice to be good examples to reveal all specific features of sub-Riemannian geometry in its generality. For instance, the cut locus and the conjugate loci for the Heisenberg group globally coincide.

A natural next step after considering nilpotent groups is to consider semisimple Lie groups. Let G be such a group with the Lie algebra $\mathfrak{g}$. Let $\iota : \mathfrak{g} \to \mathfrak{g}$ be the Cartan involution sending an element X from the Lie algebra to minus its conjugate transpose. Then there is a splitting $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$, where $\mathfrak{k}$ and $\mathfrak{p}$ are the $+1$ and -1 eigenspaces of ι. Remember that the Killing form

$$\mathrm{Kil}(X_1, X_2) = \mathrm{trace}(\mathrm{ad}_{X_1} \circ \mathrm{ad}_{X_2}), \qquad X_1, X_2 \in \mathfrak{g},$$

is non-degenerate when $\mathfrak{g}$ is semisimple. If G is compact (or more generally, if $G/\mathcal{Z}(G)$ is compact, where $\mathcal{Z}(G)$ denotes the center of G), then $\mathfrak{g} = \mathfrak{k}$, and $-\mathrm{Kil}(\cdot,\cdot)$, is positive definite on $\mathfrak{g}$. So, we can use this to define a bi-invariant Riemannian metric on G. The restriction of this metric to a distribution on G gives a sub-Riemannian manifold. An example of such a manifold, namely $\mathrm{SU}(2)$ (or S^3 considered as unit quaternions), was considered in [6, 9]. The problem of geodesic connectivity on S^3 was addressed in [9].

If G is non-compact and $\mathfrak{p} \neq 0$ (i.e., $G/\mathcal{Z}(G)$ is non-compact), then the Killing form restricted to $\mathfrak{p}$ is positive definite. We consider a left translation of $\mathfrak{p}$ as our horizontal distribution, and using the metric induced by the Killing form, we obtain a sub-Riemannian manifold equipped with a bi-invariant metric. Let K be a subgroup of G with the Lie algebra $\mathfrak{k}$. Then G is diffeomorphic to $K \times \mathfrak{p}$, by $(k, X) \mapsto ke^X$, and if G has a finite center, then K is a maximal compact subgroup. In addition, the above mentioned distribution is horizontal with respect to the quotient map

$$G \to G/K.$$

It follows that all normal geodesics are liftings of geodesics from G/K, hence they are of the form

$$t \mapsto g_0 e^{tX} e^{-\mathrm{pr}_{\mathfrak{k}}\, tX}, \qquad X \in \mathfrak{g},$$

where $g_0 \in G$ is the initial point.

Here we consider an example with $G = \mathrm{SU}(1,1)$ and $K = \mathrm{U}(1)$. Although non-holonomic geometry on $\mathrm{SU}(1,1)$ (or the isometric case of $\mathrm{SL}(2)$) was first considered earlier in, e.g., [5, 23], we will obtain new results and a more complete description of geodesics both in sub-Riemannian and sub-Lorentzian settings. Much of all meaningful results come from the analysis of the universal cover of $\mathrm{SU}(1,1)$, which we denote by $\mathrm{CSU}(1,1)$, and which is of its own interest as a new representative of a non-nilpotent Lie group over the topological space $\mathbb{R}^3$. We remark also that the Kähler manifold $\mathrm{SU}(1,1)/\mathrm{U}(1)$, is of particular importance in quantum field theories describing a black hole in two-dimensional spacetime by means of an $\mathrm{SU}(1,1)/\mathrm{U}(1)$-gauged Wess-Zumino-Novikov-Witten model, see e.g., [7, 24].

Let $\mathfrak{su}(1,1)$ be the Lie algebra of $\mathrm{SU}(1,1)$. When considered as a bilinear form on the entire $\mathfrak{su}(1,1)$, the Killing form is an index 1 pseudo-metric. Furthermore, the induced Lorentzian metric on the Lie group makes $\mathrm{SU}(1,1)$ isometric to what is called the anti-de Sitter space AdS_3 in General Relativity. This makes it tempting to study sub-Lorentzian geometry on $\mathrm{SU}(1,1)$. Apart from the fact that the Hamiltonian approach was proposed in [8] to study sub-Lorentzian structures on $\mathrm{SU}(1,1)$, the authors are not aware of other concrete examples so far, where sub-Lorentzian geometry is studied on any other manifold different from the Heisenberg group [11, 12] or an extension of it to a $\mathbb{H}$-type Carnot group [17].

The notion of distance in sub-Lorentzian geometry, as well as in Lorentzian geometry, is given by the supremum of length over timelike curves. Since timelike loops may appear in $\mathrm{SU}(1,1)$, the distance function behaves badly (more specifically the distance from a point to itself is ∞). Therefore, sub-Lorentzian geometry on $\mathrm{CSU}(1,1)$ is more interesting and meaningful than on $\mathrm{SU}(1,1)$. We are also interested in comparison of sub-Lorentzian and Lorentzian geometries on $\mathrm{CSU}(1,1)$, to try to understand somewhat more how the geometry changes when the Lorentzian metric is restricted to a distribution. The strong interplay between $\mathrm{CSU}(1,1)$ and $\mathrm{SU}(1,1)$ will, however, be practical for all explicit calculations.

The structure of the paper is as follows. Section 2 is devoted to general definitions and relations between sub-Riemannian and sub-Lorentzian geometry, and optimal control. Section 3 describes the Lie groups $\mathrm{SU}(1,1)$ and $\mathrm{CSU}(1,1)$. Section 4 contains results concerning sub-Riemannnian geometry. We describe the number of geodesics connecting two points and give explicit formulas for the distance functions. The cut and conjugate loci on both Lie groups are given. We discuss the connection between the conjugate locus and the behavior of the geodesics. In section 5, we completely describe the two-point connectivity problem by sub-Lorentzian geodesics, and compare it with the Lorentzian case. The Lorentzian and sub-Lorentzian future for $\mathrm{CSU}(1,1)$ are compared. It turns out that the reachable sets for Lorentzian and sub-Lorentzian normal geodesics intersect but are not included one to the other. Section 6 contains the proofs of main results.

The authors express special thanks to Mauricio Godoy and Irina Markina for many interesting discussions during seminars at the University of Bergen.

2. Sub Riemannian and Sub-Lorentzian geometry

2.1. **Sub-Riemannian manifolds.** A sub-Riemannian manifold is an n-dimensional manifold M, with a fiber metric ρ on an m-dimensional smooth distribution $D \to M$ ($2 \leq m \leq n$). By distribution, we mean a sub-bundle of the tangent bundle. Absolutely continuous curves that are almost everywhere tangent to D are called horizontal. The length of a horizontal curve $\gamma : [0, \tau] \to M$, is defined by

$$\ell(\gamma) := \int_0^\tau \rho^{1/2}(\dot{\gamma}(t), \dot{\gamma}(t))\, dt.$$

The *Carnot-Carathéodory distance* between points $q_0, q_1 \in M$ is defined as

$$d(q_0; q_1) = \inf_{\gamma} \ell(\gamma),$$

where the infimum is taken over all horizontal curves γ satisfying $\gamma(0) = q_0$ and $\gamma(\tau) = q_1$. If there are no such curves connecting q_0 and q_1, then the distance is ∞. If the minimum in the above relation is attained, then the curve is called a length minimizing curve.

Define $D^1 = D$ and iteratively $D^i = D^{i-1} + [D, D^{i-1}]$, $i \geq 2$, where $[\cdot, \cdot]$ denotes the Lie brackets, $[X, Y] = XY - YX$ for $X, Y \in T_qM$, $q \in M$. If there exists a positive integer $k \geq 2$, such that $D^k = TM$, then the distribution is called bracket generating. It is called regular if $\dim D^i_q$ is independent of the choice of q for all i. We say that D is step k regular, if it is regular and k is the smallest number for which $D^k = TM$. Chow-Rashevskiĭ theorem [20, 10] states that a bracket generating distribution D guarantees that any two points may be connected by a horizontal curve. In addition, we have the following generalizations of the corresponding properties from the Riemannian case:

Theorem 1 (Hopf-Rinow Theorem for sub-Riemannian manifolds [3])**.** *Suppose that D satisfies the bracket generating condition. Then*

i) *any $q_0 \in M$ has a neighborhood U, such that there exists a minimizing curve joining the points q_1 and q_0 for every $q_1 \in U$;*
ii) *if M is complete regarding to d, then any two points can be joined by a minimizing curve.*

Remark that the length minimizing curve may be singular for arbitrarily close points (see [18]).

A curve is called geodesic if it is locally a length minimizneighborhooder. By a normal geodesic in the sub-Riemannian case we mean an integral curve of the Hamiltonian system generated by a Hamiltonian function $H = \sum_{j=1}^m h_j^2$ in some neighborhood of a point q with respect to any local orthonormal basis $\{X_1, \dots, X_m\}$ in this neighborhood. Here, and in rest of the paper, if X is a vector field, then $h_X : T^*M \to \mathbb{R}$ denotes the Hamiltonian function

$$h_X : \lambda \mapsto \lambda(X(q)), \qquad \lambda \in T_qM,$$

with respect to the vector field X. If we have a basis $\{X_1, \dots, X_n\}$, then we simplify notations by writing h_j instead of h_{X_j}. One of the principal differences between sub-Riemannian

and Riemannian geometries is that the function $q \mapsto d(q_0; q)$ is not differentiable in any neighborhood of q_0 when $m < n$.

2.2. **Sub-Lorentzian manifolds.** A sub-Lorentzian manifold is defined similarly to a sub-Riemannian manifold, but with ρ now being an index 1 pseudo-metric on D. We will say that a vector $v \in D$ is

- timelike if $\rho(v, v) < 0$,
- lightlike or null if $\rho(v, v) = 0$,
- spacelike if $\rho(v, v) > 0$,
- causal or nonspacelike if $\rho(v, v) \leq 0$.

A chosen vector field T in D, is said to be the time-orientation on M, if $\rho(T(q), T(q)) < 0$ for any $q \in M$. A causal vector is called future directed if $\rho(T(q), v) < 0$, and past directed if $\rho(T(q), v) > 0$. A horizontal curve $\gamma : [0, \tau] \to \mathbb{R}$ is called timelike, null, spacelike, causal, future directed or past directed, respectively, if $\dot{\gamma}(t)$ is such a vector for almost every $t \in [0, \tau]$. We define the timelike future $\mathcal{I}^+(q_0, \rho)$ of q_0 with respect to ρ as the set of all points $q_1 \in M$, such that there is a horizontal, timelike future directed curve γ, with $\gamma(0) = q_0$ and $\gamma(\tau) = q_1$. The causal future, $\mathcal{J}^+(q_0, \rho)$, is defined similarly, with timelikeness interchanged with causality. Analogous definitions are valid for the timelike or causal past, which we denote by $\mathcal{I}^-(q_0, \rho)$ and $\mathcal{J}^-(q_0, \rho)$ respectively. We define the length of a horizontal causal curve by $\ell(\gamma) = \int_0^\tau |\rho(\dot{\gamma}(t), \dot{\gamma}(t))|^{1/2} dt$. The sub-Lorentzian distance is defined by

$$d(q_0; q_1) = \begin{cases} \sup\limits_{\gamma} \ell(\gamma), & \text{if } q_1 \in \mathcal{J}^+(q_0, \rho), \\ 0, & \text{otherwise.} \end{cases}$$

The supremum is taken over all horizontal future directed causal curves from q_0 to q_1. Similarly to the Lorentzian distance, the sub-Lorentzian distance satisfies the reverse triangle inequality, and may not be very well behaving. For instance, if there is a timelike loop trough a point $q \in M$, then $d(q; q) = \infty$.

A curve $\gamma : [0, \tau] \to \mathbb{R}$ is called a length maximizer, if $\ell(\gamma) = d(\gamma(0); \gamma(\tau))$. Similarly, a curve γ is called a relative maximizer with respect to an open set W, if $\gamma([0, \tau]) \subseteq W$ and $\ell(\gamma) = \sup_{\tilde{\gamma}} \ell(\tilde{\gamma})$, where the supremum is taken over all horizontal future directed causal curves contained in W, connecting $\gamma(0)$ and $\gamma(\tau)$. By using the maximum principle, for D bracket generating, we know that all relative maximizers are either normal geodesics or strictly abnormal maximizers [12], and that the relative maximizers always exist locally. By normal sub-Lorentzian geodesics γ we mean curves, such that for any local orthonormal basis $\{X_1, \dots, X_m\}$ of D, with X_1 as the time-orientation, γ is an integral curve of the Hamiltonian system generated by a Hamiltonian function $-h_1^2 + \sum_{j=2}^m h_j^2$.

The question of whether length maximizers exist between two points, is a much more complicated than the question of the existence of length minimizers in Riemannian or sub-Riemannian geometries. The most common sufficient condition for the global existence of maximizing curves on a Lorentzian or sub-Lorentzian manifold M is a relatively strict requirement that M should be globally hyperbolic, i.e., strongly causal (every point has an

arbitrarily small neighborhood, such that causal curves that leave the neighborhood never return back), and $\mathcal{J}^+(q_1,\rho)\cap\mathcal{J}^-(q_2,\rho)$ is compact for any $q_1,q_2\in M$.

2.3. **Minimizing and maximizing curves seen from the viewpoint of optimal control.** Determination of curves whose length is equal to the distance, either in the sub-Riemannian or sub-Lorentzian setting, can be formulated as a solution to an optimal control problem. Let M be an n-dimensional manifold, let a submersion $\pi : V\to M$ be a fiber bundle with the fiber $U=\pi^{-1}(q)$, $q\in M$, and let $f : V\to TM$ be a fiber preserving map.

An *admissible pair* $v=(u,\gamma_u)$ is a bounded measurable mapping $v:[0,\tau]\to V, 0<\tau<\infty$ such that $\gamma_u=\pi\circ v$ is a Lipschitzian curve on M and $\dot\gamma_u=f(u,\gamma_u)$. For some fixed value of τ, we denote the space of all admissible pairs by $\mathcal{V}_\tau$. The space $\mathcal{V}_\tau$ is a smooth Banach submanifold of $L^\infty([0,\tau],V)$, which is modeled on $L^\infty([0,\tau],\mathbb{R}^{\dim V})$. An admissible pair is said to connect $q_0,q_1\in M$, if $\gamma_u(0)=q_0$, $\gamma_u(\tau)=q_1$. Such a system is called *controllable*, if any two points in M are connected by at least one admissible pair. Let $J:\mathcal{V}_\tau\to\mathbb{R}$ be a functional, given by

$$J(v)=\int_0^\tau K(v)dt\,,$$

where K is some differentiable function. Suppose we are given two points $q_0,q_1\in M$, such that there is at least one admissible pair connecting them. An *optimal control problem* with respect to the functional J, is a problem of finding an element $v^*=(u^*,\gamma_{u^*})\in\mathcal{V}_\tau$ connecting q_0,q_1, such that for any $v\in\mathcal{V}_\tau$ connecting the same two points, we have $J(v^*)\le J(v)$. Analogously, we may look for the maximum of J. We call u^* the optimal control, and γ_{u^*} the optimal trajectory. We may also consider *free-time* optimal control problems where τ is allowed to vary.

The main tool to solve optimal control problems is the following first order condition known as the *Pontryagin Maximum Principle* (PMP).

Theorem 2 (PMP for Optimal Control Problem)**.** *For a given value of τ, let $v^*=(u^*,\gamma_{u^*})\in\mathcal{V}_\tau$ be an optimal pair for the above problem, i.e., $J(v^*)=\min\limits_{v\in\mathcal{V}_\tau}\{J(v)\big|\,\gamma_u(0)=0,\gamma_u(\tau)=q_1\}$. Let us define a pseudo-Hamiltonian function as*

$$\mathscr{H}_\varkappa(u,\lambda)=\lambda(f(u,q))+\varkappa K(u,q),\qquad \lambda\in T_q^*M,\quad u\in U,\quad \varkappa\in\mathbb{R}\,.$$

Then there exists a curve $\xi:[0,\tau]\to T^*M$, and a number $\varkappa\le 0$, such that

$$\dot\xi(t)=\vec{\mathscr{H}}_{\varkappa,}(u^*,\xi(t)),\qquad \mathscr{H}_\varkappa(u^*,\xi(t))=\max_{u\in U}\mathscr{H}_\varkappa(u,\xi(t))\quad a.e.\ in\ t\in[0,\tau],$$

and

$$\mathrm{pr}_M\,\xi(t)=\gamma_{u^*}.$$

Moreover, if $\varkappa=0$, then $\xi(t)$ does not vanish almost everywhere in $t\in[0,\tau]$.

We will write $\vec H$ to denote the Hamiltonian vector field associated to a Hamiltonian function H. For a pseudo-Hamiltonian function, $\vec{\mathscr{H}}$ is defined so that for a fixed $u\in U$, $\vec{\mathscr{H}}(u,\cdot)$ is the Hamiltonian vector field associated with the Hamiltonian $\mathscr{H}(u,\cdot)$. For the

problem of the maximum of J, the above theorem has the same formulation, changing only the requirement $\varkappa \leq 0$ to $\varkappa \geq 0$. If we consider a free-time problem, then we also require $\mathscr{H}_{\varkappa}(u^*, \xi(t)) \equiv 0$. If $\varkappa \neq 0$, then the solution is called normal (in this case we may just choose $\varkappa = \pm 1$). If $\varkappa = 0$, the solution is called abnormal.

Remark 1 ([2]). *If $H_{\varkappa}(\lambda) = \max_{u \in U} \mathscr{H}_{\varkappa}(u, \lambda)$ is defined and is C^2 on $T^*M \setminus s_0(M)$, where $s_0 : M \to T^*M$ is the zero-section, then $\dot{\xi}(t) = \vec{H}_{\varkappa}(\xi(t))$.*

Let us turn to specific sub-Riemannian and sub-Lorentzian settings. We only need the formulation for the case when $M = G$ is a Lie group. This case is somewhat simpler since the tangent bundle of Lie groups is trivial, so we can always find a global basis to span a distribution. Let $D = \operatorname{span}\{X_1, \ldots, X_m\}$ be a distribution, and ρ be either a sub-Riemannian or sub-Lorentzian metric on D. We may assume that the collection $\{X_j\}_{j=1}^m$ forms an orthonormal basis. Let $V = G \times U$, where $U \subset \mathbb{R}^m$. We choose $f(u_1, \ldots, u_m, g) = u_1 X_1(g) + \cdots + u_m X_m(g)$, $g \in G$, to ensure that $g(t)$ is a horizontal curve.

The determination of the sub-Riemannian distance, comes down to finding an optimal pair $(u^*, g_{u^*}) : [0, \tau] \to U \times G$ that minimizes the functional

$$J(u, g_u) = \frac{1}{2} \int_0^\tau \sum_{j=1}^m u_j^2 \, dt, \qquad u \in U = \{u \in \mathbb{R}^m | \, \|u\| \leq 1\}.$$

The corresponding pseudo-Hamiltonian is given by

$$\mathscr{H}_\kappa(u, \lambda) = -\frac{\varkappa}{2} \sum_{j=1}^m u_j^2 + \sum_{j=1}^m u_k h_k(\lambda), \quad \lambda \in T^*G, \quad u \in U, \quad \varkappa = 0, 1.$$

As a consequence of Fillipov's theorem [2], there always exists a length minimizing horizontal curve between two points in this setting.

Similarly, finding the sub-Lorentzian distance with X_1 as a time-orientation, means finding a pair (u^*, g_{u^*}) that maximizes

$$J(u, g_u) = \int_0^\tau \left(u_1^2 - \sum_{j=2}^m u_j^2 \right)^{\frac{1}{2}} dt, \qquad U = \left\{ u \in \mathbb{R}^m \Big| \, u_1 = \left(1 + \sum_{j=2}^m u_j^2 \right)^{\frac{1}{2}} \right\},$$

and the pseudo-Hamiltonian related to this problem becomes

$$\mathscr{H}_\kappa(u, \lambda) = \varkappa \left(u_1^2 - \sum_{j=2}^m u_j^2 \right)^{\frac{1}{2}} + \sum_{j=1}^m u_k h_k(\lambda).$$

The latter case is more complicated because we optimize a non-convex functional.

In our example of $\mathrm{SU}(1,1)$, we only consider left-invariant distributions, that lead to left-invariant Hamiltonians. Let us, therefore, include some sketch of the theory of Hamiltonian systems in this case. Generally, assume that H is a left-invariant Hamiltonian function on a Lie group (i.e., $H(d^* L_g \lambda) = H(\lambda)$). From the isomorphisms of bundles, we have

$$G \times \mathfrak{g} \to TG, \qquad G \times \mathfrak{g}^* \to T^*G,$$

$$(g, X) \mapsto dL_g X, \qquad (g, p) \mapsto d^* L_{g^{-1}} p.$$

The traditional Hamiltonian equation $\dot{\lambda}(t) = \vec{H}(\lambda(t))$, which locally has the form

$$\dot{g} = \frac{\partial H}{\partial \xi}, \qquad \dot{\xi} = -\frac{\partial H}{\partial g}, \qquad \lambda = (g, \xi) \in T^* M,$$

may be written as

$$\dot{g} = dL_g \left(\frac{\partial H}{\partial p}(g, p) \right),$$

$$\dot{p} = d^* L_g \left(\frac{\partial H}{\partial g}(g, p) \right) + \operatorname{ad} \left(\frac{\partial H}{\partial p}(g, p) \right)^* (p).$$

If H is left-invariant, and hence is independent of g, then the first term in the equation for $\dot{p}$ vanishes. In particular, if X is a left-invariant vector field, then $h_X(g, p) = d^* L_{g^{-1}} p(X(g)) = p(dL_{g^{-1}} dL_g X(1)) = p(X(1))$, so $h_j(p) = p(X_j) =: p_j$. Furthermore,

$$\begin{aligned} \dot{p}_j &= \dot{p}(X_j) = \operatorname{ad} \left(\frac{dH}{dp}(p) \right)^* (p)(X_j) = p \left(\left[\frac{dH}{dp}(p), X_j \right] \right) \\ &= p \left(\left[\frac{dH}{dp}(p), \frac{dh_j}{dp}(p) \right] \right) = \{h_j, H\}(p). \end{aligned}$$

Here $\{\cdot, \cdot\}$ denotes the Poisson brackets. For more details, see [15] (observe that the difference in sign in our formulation and in [15] happens because of different definitions of $[\cdot, \cdot]$.) Notice that similar considerations can be done by interchanging the left and the right actions.

3. Structure of the Lie groups SU(1,1) and CSU(1,1)

3.1. **Some notation.** We will start by defining the following functions, which will simplify our notations later. First, let $\alpha : \mathbb{R}^3 \mapsto \mathbb{R}$ be defined by

$$\alpha : a = (a_1, a_2, a_3) \mapsto \alpha_a = a_1^2 + a_2^2 - a_3^2.$$

Furthermore, let

$$C(\alpha, t) = \begin{cases} \cosh(\sqrt{\alpha} \cdot t), & \text{if } \ \alpha \geq 0, \\ \cos(\sqrt{-\alpha} \cdot t), & \text{if } \ \alpha < 0, \end{cases} \qquad S(\alpha, t) = \begin{cases} \dfrac{\sinh(\sqrt{\alpha} \cdot t)}{\sqrt{\alpha}}, & \text{if } \ \alpha > 0, \\ t & \text{if } \ \alpha = 0, \\ \dfrac{\sin(\sqrt{-\alpha} \cdot t)}{\sqrt{-\alpha}} & \text{if } \ \alpha < 0, \end{cases}$$

where $(\alpha, t) \in \mathbb{R} \times \mathbb{R}_+$. We remark that

$$\frac{\partial}{\partial t} C(\alpha, t) = \alpha S(\alpha, t), \qquad \frac{\partial}{\partial t} S(\alpha, t) = C(\alpha, t),$$

$$\frac{\partial}{\partial \alpha} C(\alpha, t) = \frac{t}{2} S(\alpha, t), \qquad \frac{\partial}{\partial \alpha} S(\alpha, t) = \frac{t}{2\alpha} \left(C(\alpha, t) - S(\alpha, t) \right).$$

Also

$$C^2(\alpha, t) - \alpha S^2(\alpha, t) = 1.$$

Observe that $\frac{\partial}{\partial\alpha}S(\alpha,t)|_{\alpha=0}$ does not exist, because

$$\lim_{\alpha\to 0+}\frac{\partial}{\partial\alpha}S(\alpha,t)=\frac{t}{6},\quad\text{while}\quad\lim_{\alpha\to 0-}\frac{\partial}{\partial\alpha}S(\alpha,t)=-\frac{t}{6}.$$

When $t=1$, we will write $C(\alpha):=C(\alpha,1)$ and $S(\alpha):=S(\alpha,1)$.

Remark 2. *Define $\chi_k, k\in\mathbb{N}$, as a unique number satisfying $\chi_k=\tan\chi_k, \pi k\le\chi_k\le \pi k+\frac{\pi}{2}$. Then, $\frac{\partial}{\partial\alpha}C(\alpha)=0$, implies that $\alpha=-\pi^2k^2$, while $\frac{\partial}{\partial\alpha}S(\alpha)=0$ implies $\alpha=-\chi_k^2$. These numbers will become important later.*

Finally, we define $\phi:\mathbb{R}^3\to\mathbb{R}$ in the following way. If $\alpha_a\ge 0$, then $\phi(a)=\tan^{-1}\left(\frac{a_3S(\alpha_a)}{C(\alpha_a)}\right)$, that is

$$\phi(a)=\tan^{-1}\left(\frac{a_3}{\sqrt{\alpha_a}}\tanh\sqrt{\alpha_a}\right),$$

when $\alpha_a>0$, and $\phi(a)=\tan^{-1}a_3$ when $\alpha_a=0$. Furthermore, if $\alpha_a<0$, then

$$\phi(a)=\begin{cases}\tan^{-1}\left(\frac{a_3}{\sqrt{-\alpha_a}}\tan\sqrt{-\alpha_a}\right)+\operatorname{sgn}(a_3)\pi\left\lceil\frac{\sqrt{-\alpha_a}}{\pi}-\frac{1}{2}\right\rceil, & \text{if } \sqrt{-\alpha_a}\ne\pi/2 \bmod \pi,\\ \operatorname{sgn}(a_3)\sqrt{-\alpha_a}, & \text{if } \sqrt{-\alpha_a}=\pi/2 \bmod \pi.\end{cases}$$

Here, $\lceil x\rceil$ denotes the ceiling of x (i.e., $\lceil x\rceil=\min\{j\in\mathbb{Z}|j\ge x\}$) and

$$\operatorname{sgn}x=\begin{cases}1, & \text{if } x>0,\\ 0, & \text{if } x=0,\\ -1, & \text{if } x<0.\end{cases}$$

3.2. **Lie group** $\mathrm{SU}(1,1)$**.** We consider the Lie group $\mathrm{SU}(1,1)$ of unitary complex (2×2) matrices of the form

$$g=\begin{pmatrix} z_1 & z_2\\ \bar z_2 & \bar z_1\end{pmatrix},$$

with the operation of usual matrix multiplication. Simplifying notation, we will sometimes view $g\in\mathrm{SU}(1,1)$ as an element from $\mathbb{C}^2$, writing $g=(z_1,z_2)$. In this notation, the group operation is

$$(z_1,z_2)(\hat z_1,\hat z_2)=\left(z_1\hat z_1+z_2\overline{\hat z_2}, z_1\hat z_2+z_2\overline{\hat z_1}\right),$$

the identity is $1=(1,0)$, and the inverse element becomes $(z_1,z_2)^{-1}=(\bar z_1,-z_2)$. $\mathrm{SU}(1,1)$ is isomorphic to the Lie group $\mathrm{SL}(2)$ of real 2×2 matrices of determinant 1, by the mapping[1]

$$\begin{array}{ccc}\mathrm{SU}(1,1) & \cong & \mathrm{SL}(2),\\ \begin{pmatrix} z_1 & z_2\\ \bar z_2 & \bar z_1\end{pmatrix} & \mapsto & \begin{pmatrix}\mathrm{Re}\ z_1+\mathrm{Im}\ z_2 & ; & -\mathrm{Im}\ z_1-\mathrm{Re}\ z_2\\ \mathrm{Im}\ z_1-\mathrm{Re}\ z_2 & ; & \mathrm{Re}\ z_1-\mathrm{Im}\ z_2\end{pmatrix}.\end{array}$$

We will often use the isomorphism from the tangent bundle $T\,\mathrm{SU}(1,1)$ to a subspace of the complex tangent bundle $T_{\mathbb{C}}\,\mathrm{SU}(1,1)$, given by

$$u_1\partial_{x_1}+u_2\partial_{y_2}+u_3\partial_{y_3}+u_4\partial_{x_4}\mapsto(u_1+iu_2)\partial_{z_1}+(u_1-iu_2)\partial_{\bar z_1}+(u_3+iu_4)\partial_{z_2}+(u_3-iu_4)\partial_{\bar z_2},$$

[1] We shall place the semicolon (;) in some matrices in order to separate long expressions.

where $z_j = x_j + iy_j$, $\partial_{z_j} = \frac{1}{2}(\partial_{x_j} - i\partial_{y_j})$ and $\partial_{\bar{z}_j} = \frac{1}{2}(\partial_{x_j} + i\partial_{y_j})$. Let L_g denote the left actions by g. The tangent map of the left action is given by

$$dL_g : b_1\partial_{z_1} + \bar{b}_1\partial_{\bar{z}_1} + b_2\partial_{z_2} + \bar{b}_2\partial_{\bar{z}_1} \mapsto$$

$$\mapsto (b_1 z_1 + \bar{b}_2 z_2)\partial_{z_1} + (\bar{b}_1\bar{z}_1 + b_2\bar{z}_2)\partial_{\bar{z}_1} + (b_2 z_1 + \bar{b}_1 z_2)\partial_{z_2} + (\bar{b}_2\bar{z}_1 + b_1\bar{z}_2)\partial_{\bar{z}_2}.$$

The tangent space at the identity is spanned by the vectors

$$X(1) = -\partial_{z_2} - \partial_{\bar{z}_2}, \quad Y(1) = i\partial_{z_2} - i\partial_{\bar{z}_2}, \quad Z(1) = -i\partial_{z_1} + i\partial_{\bar{z}_1}.$$

Viewed as elements of the Lie algebra $\mathfrak{su}(1,1)$ of $\mathrm{SU}(1,1)$, they have the following form

$$X = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix},$$

analogously to the Pauli matrices in relation to the Lie algebra $\mathfrak{su}(2)$ of the Lie group $\mathrm{SU}(2)$. So we can obtain the left-invariant basis for the tangent bundle as

$$\begin{aligned} X(g) &= -z_2\partial_{z_1} - \bar{z}_2\partial_{\bar{z}_1} - z_1\partial_{z_2} - \bar{z}_1\partial_{\bar{z}_2}, \\ Y(g) &= -iz_2\partial_{z_1} + i\bar{z}_2\partial_{\bar{z}_1} + iz_1\partial_{z_2} - i\bar{z}_1\partial_{\bar{z}_2}, \\ Z(g) &= -iz_1\partial_{z_1} + i\bar{z}_1\partial_{\bar{z}_1} + iz_2\partial_{z_2} - i\bar{z}_2\partial_{\bar{z}_2}. \end{aligned}$$

The dual basis for the cotangent bundle is

$$X^*(g) = \frac{1}{2}(\bar{z}_2 dz_1 + z_2 d\bar{z}_1 - \bar{z}_1 dz_2 - z_1 d\bar{z}_2),$$

$$Y^*(g) = \frac{1}{2}(-i\bar{z}_2 dz_1 + iz_2 d\bar{z}_1 - i\bar{z}_1 dz_2 + iz_1 d\bar{z}_2),$$

$$Z^*(g) = \frac{1}{2}(i\bar{z}_1 dz_1 - iz_1 d\bar{z}_1 + i\bar{z}_2 dz_2 - iz_2 d\bar{z}_2).$$

The bracket relations yield

$$[X, Y] = -2Z, \qquad [X, Z] = -2Y, \qquad [Y, Z] = 2X.$$

It follows that any distribution spanned by two of three vector fields is bracket generating.

The exponential map for this Lie group is

$$e^{a_1X + a_2Y + a_3Z} = \begin{pmatrix} C(\alpha_a) - ia_3S(\alpha_a) & ; & -(a_1 - ia_2)S(\alpha_a) \\ -(a_1 + ia_2)S(\alpha_a) & ; & C(\alpha_a) + ia_3S(\alpha_a) \end{pmatrix}$$

$$= C(\alpha_a) \cdot 1 + S(\alpha_a)(a_1X + a_2Y + a_3Z).$$

We define the metric ρ as a left-invariant metric, whose restriction to the $\mathfrak{su}(1,1)$ has the following form

$$\rho : (X_1, X_2) \mapsto \frac{1}{8}\mathrm{Kil}(X_1, X_2)$$

Remark that in $\mathfrak{su}(1,1)$, the Killing form is equal to

$$\mathrm{Kil}(X_1, X_2) = 4\,\mathrm{trace}(X_1X_2).$$

In the basis of $\partial_{x_1}, \partial_{y_1}, \partial_{x_2}, \partial_{y_2}$, the metric tensor of ρ has the form

$$(\rho_{ij}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The metric ρ is then Lorentzian, and in fact, bi-invariant. Hence, as a Lorentzian manifold, $\mathrm{SU}(1,1)$ may be considered as a subset of $\mathbb{R}^{2,2}$, which is $\mathbb{R}^4$ with an index 2 pseudo-metric. This subset

$$\mathrm{AdS}_3 = \left\{(x_1, y_1, x_2, y_2) \in \mathbb{R}^{2,2} \,|\, x_1^2 + y_1^2 - x_2^2 - y_2^2 = 1\right\}$$

is called the 3-dimensional anti-de Sitter space.

The restriction of ρ to the distribution $D = \mathrm{span}\{X, Y\}$, makes it positive definite, and makes $(\mathrm{SU}(1,1), D, \rho|_D)$ a sub-Riemannian manifold. Similarly, the restriction ρ to $E = \mathrm{span}\{Y, Z\}$ makes $(\mathrm{SU}(1,1), E, \rho|_E)$ a sub-Lorentzian manifold, if we define the time-orientation by Z. The latter case however, contains timelike loops. An example is

$$t \mapsto \begin{pmatrix} e^{-it} & 0 \\ 0 & e^{it} \end{pmatrix},$$

which is a loop through 1. To avoid this problem we will also study the universal cover $\mathrm{CSU}(1,1)$ of $\mathrm{SU}(1,1)$. This will also be helpful in order to study sub-Riemannian geometry on $\mathrm{SU}(1,1)$. In addition, it is an interesting example on its own. In fact, sometimes the attribution anti-de Sitter space is used for the universal cover instead of AdS_3 itself.

3.3. **Lie group structure of the universal cover $\mathrm{CSU}(1,1)$ of $\mathrm{SU}(1,1)$.** Since $\mathrm{SU}(1,1)$ is diffeomorphic to $S^1 \times \mathbb{R}^2$, the universal cover must be diffeomorphic to $\mathbb{R}^3$. We represent the covering space $\mathrm{CSU}(1,1)$ as $\mathbb{R} \times \mathbb{C}$, with the covering map

$$\pi : \mathrm{CSU}(1,1) \to \mathrm{SU}(1,1),$$

$$\pi : \tilde{g} = (c, w) \mapsto \begin{pmatrix} \sqrt{1 + |w|^2} e^{ic} & ; & w \\ \bar{w} & ; & \sqrt{1 + |w|^2} e^{-ic} \end{pmatrix},$$

where $c \in \mathbb{R}$ and $w \in \mathbb{C}$. We define the product on $\mathrm{CSU}(1,1)$, to be the unique product for which $\tilde{1} = (0, 0)$ is the identity, and which makes π a group homomorphism. It is obvious that the Lie algebra of $\mathrm{CSU}(1,1)$ is also $\mathfrak{su}(1,1)$.

Proposition 1. *Let $(c_j, w_j) \in \mathrm{CSU}(1,1),\ j = 1, 2$. Then* $(c_1, w_1)(c_2, w_2) = (c, w)$, *where*

$$c = c_1 + c_2 + \tan^{-1}\left(\frac{\mathrm{Im}\ (w_1 \bar{w}_2 e^{-i(c_1+c_2)})}{\sqrt{(1 + |w_1|^2)(1 + |w_2|^2)} + \mathrm{Re}\ (w_1 \bar{w}_2 e^{-i(c_1+c_2)})}\right),$$

$$w = w_2 \sqrt{1 + |w_1|^2} e^{ic_1} + w_1 \sqrt{1 + |w_2|^2} e^{-ic_2}.$$

Proof. The fact that $\pi(c_1, w_1)\pi(c_2, w_2) = \pi(c, w)$, trivially implies the expression for w. The value of c must satisfy the relation

$$e^{ic} = \exp\left(i\mathrm{Arg}\left(\sqrt{(1+|w_1|^2)(1+|w_2|^2)}e^{i(c_1+c_2)} + w_1\bar{w}_2\right)\right)$$
$$= e^{i(c_1+c_2)} \exp\left(i\mathrm{Arg}\left(\sqrt{(1+|w_1|^2)(1+|w_2|^2)} + w_1\bar{w}_2 e^{-i(c_1+c_2)}\right)\right).$$

Now since

$$\left|\sqrt{(1+|w_1|^2)(1+|w_2|^2)}\right| > \left|w_1\bar{w}_2 e^{-i(c_1+c_2)}\right|,$$

we know that

$$-\frac{\pi}{2} < \mathrm{Arg}\left(\sqrt{(1+|w_1|^2)(1+|w_2|^2)} + w_1\bar{w}_2 e^{-i(c_1+c_2)}\right) < \frac{\pi}{2},$$

and the formula for c follows. It is clear that $(0,0)$ is the identity under this product. Observe that $(c,w)^{-1} = (-c,-w)$. The associativity of the product remains to be proved. Due to the associativity of the product on $\mathrm{SU}(1,1)$, we only need to show that if

$$(C, W) = ((c_1, w_1)(c_2, w_2))\,(c_3, w_3), \qquad (\widehat{C}, \widehat{W}) = (c_1, w_1)\,((c_2, w_2)(c_3, w_3)),$$

then $0 = C - \widehat{C}$. Let $\pi(C, W) = \pi(\widehat{C}, \widehat{W}) = (Z_1, Z_2)$, and denote

$$\theta = \mathrm{Arg}\left(\sqrt{(1+|w_1|^2)(1+|w_2|^2} + w_1\bar{w}_2 e^{-i(c_1+c_2)}\right),$$
$$\widehat{\theta} = \mathrm{Arg}\left(\sqrt{(1+|w_2|^2)(1+|w_3|^2} + w_2\bar{w}_3 e^{-i(c_1+c_2)}\right).$$

Then

$$C = c_1 + c_2 + c_3 + \theta + \mathrm{Arg}(Z_1 e^{-i(c_1+c_2+c_3+\theta)}), \quad \widehat{C} = c_1 + c_2 + c_3 + \widehat{\theta} + \mathrm{Arg}(Z_1 e^{-i(c_1+c_2+c_3+\hat{\theta})}).$$

Similarly to the reasoning above,

$$\mathrm{Arg}(Z_1 e^{-i(c_1+c_2+c_3+\theta)}) \text{ and } \mathrm{Arg}(Z_1 e^{-i(c_1+c_2+c_3+\widehat{\theta})}) \in (-\frac{\pi}{2}, \frac{\pi}{2}),$$

so since $|\theta - \widehat{\theta}| < \pi$, we know that $\mathrm{Arg}(Z_1 e^{-i(c_1+c_2+c_3+\theta)}) - \mathrm{Arg}(Z_1 e^{-i(c_1+c_2+c_3+\hat{\theta})}) = \widehat{\theta} - \theta$, and it follows that $C = \widehat{C}$. ∎

The left action is given by ($b_0 \in \mathbb{R}, b \in \mathbb{C}$)

$$dL_{\tilde{g}} : b_0\partial_c + b\partial_w + \bar{b}\partial_{\bar{w}} \mapsto$$
$$(b_0 + \mathrm{Im}\ (\bar{b}we^{-ic}))\partial_c + (-ib_0 w + b\sqrt{1+|w|^2}e^{ic})\partial_w + (ib_0\bar{w} + \bar{b}\sqrt{1+|w|^2}e^{-ic})\partial_{\bar{w}}$$
$$= \left(b_0 + \mathrm{Im}\ \left(\frac{\bar{b}z_2|z_1|}{z_1}\right)\right)\partial_c + (-ib_0 z_2 + bz_1)\partial_w + (ib_0\bar{z}_2 + \bar{b}\bar{z}_1)\partial_{\bar{w}}.$$

The lifted vector fields are

$$\widetilde{X}(c, w) = -\mathrm{Im}\ (we^{ic})\partial_c + \sqrt{1+|w|^2}(e^{ic}\partial_w + e^{-ic}\partial_{\bar{w}}),$$
$$\widetilde{Y}(c, w) = -\mathrm{Re}\ (we^{ic})\partial_c - i\sqrt{1+|w|^2}(e^{ic}\partial_w - e^{-ic}\partial_{\bar{w}}),$$
$$\widetilde{Z}(c, w) = -\partial_c + iw\partial_w - i\bar{w}\partial_{\bar{w}}.$$

The exponential map in $\mathrm{CSU}(1,1)$ is

$$e^{a_1X+a_2Y+a_3Z}=\big(-\phi(a),-(a_1-ia_2)S(\alpha_a)\big),$$

which can be found by lifting the exponential map from $\mathrm{SU}(1,1)$.

We lift the metric ρ to a metric on $\mathrm{CSU}(1,1)$. Denote this metric by $\tilde{\rho}$. Analogously to $\mathrm{SU}(1,1)$, let us define the distributions $\widetilde{D}=\mathrm{span}\{\widetilde{X},\widetilde{Y}\}$ and $\tilde{E}=\mathrm{span}\{\widetilde{Y},\widetilde{Z}\}$. The restriction of $\tilde{\rho}$ to $\widetilde{D}$ and $\widetilde{E}$, defines sub-Riemannian and sub-Lorentzian structures on $\mathrm{CSU}(1,1)$ respectively. In the sub-Lorentzian case, we let $\widetilde{Z}$ define the time orientation.

Remark 3. *We can construct a sub-Lorentzian manifold, by considering the distribution* $\mathrm{span}\{\widetilde{X},\widetilde{Z}\}$*, but the geodesics are very similar to the sub-Lorentzian manifold* $(\mathrm{CSU}(1,1),\tilde{E},\widetilde{\rho}|_{\tilde{E}})$ *so we omit this choice of distribution.*

4. Sub-Riemannian geometry on SU(1,1) and CSU(1,1)

4.1. **Geodesics, horizontal space, and vertical space.** We will now take advantage of the fact that the pseudo-metric induced by the Killing form is bi-invariant.

Theorem 3. *Let G be a Lie group with the Lie algebra $\mathfrak{g}$, and with a bi-invariant pseudo-metric ρ. Let K be a subgroup of G, with the Lie algebra $\mathfrak{k}$, and let us denote $\mathfrak{p}=\mathfrak{k}^{\perp}$. Define a left-invariant distribution D by $D_g=dL_g\mathfrak{p}$. Then all normal geodesics on the non-holonomic manifold $(G,D,\rho|_D)$ are lifting of the normal geodesics on G/K with the induced metric. This means that all normal geodesics starting at $g_0\in G$ are of the form*

$$t\mapsto g_0e^{Xt}\cdot e^{-\mathrm{pr}_{\mathfrak{k}}Xt},\qquad X\in\mathfrak{g}\,,$$

where $\mathrm{pr}_{\mathfrak{k}}:\mathfrak{g}\to\mathfrak{k}$ is the projection.

This theorem is a special case of the corresponding result from [19].

We may use Theorem 3 with $K=e^{\mathrm{span}\{Z\}}$. The normal geodesics starting at the identity admit the form

$$z_1(t)=(C(\alpha_a,t)-ia_3S(\alpha_a,t))e^{ia_3t},\qquad z_2(t)=-(a_1-ia_2)S(\alpha_a,t)e^{-ia_3t},$$

in $\mathrm{SU}(1,1)$ and

$$c(t)=-\phi(ta)+a_3t,\qquad w(t)=-(a_1-ia_2)S(\alpha_a,t)e^{-ia_3t},$$

in $\mathrm{CSU}(1,1)$. Any other normal geodesic starting at some point, is a left translation of a normal one starting at the identity.

We define *the vertical space* in a sub-Riemannian manifold (M,D,ρ) with respect to a point q, as the set of all points in M that can be reached from q by curves tangent to the orthogonal complement of D. We define *the horizontal space* with respect to q, to be the set of all point that can be reached by horizontal curves of constant speed (that is, γ is a horizontal curve, such that if any horizontal vector field X is of unit length, then $X^*(\dot{\gamma})$ is constant).

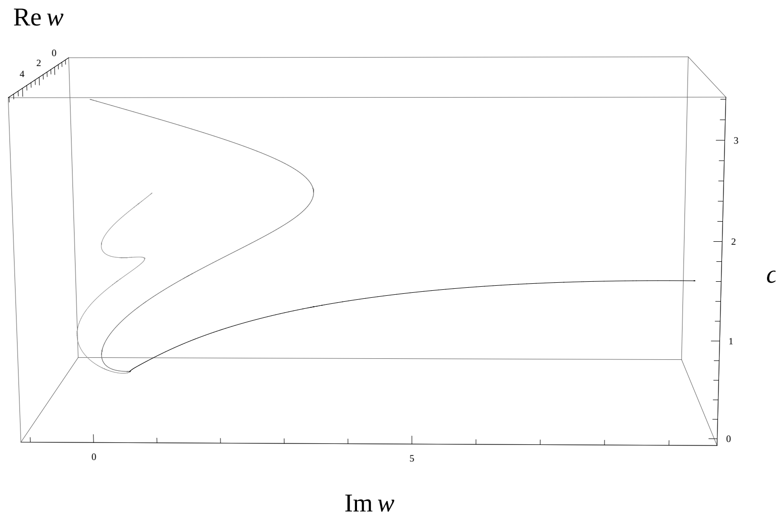

FIGURE 1. The above figure shows three geodesics in $\mathrm{CSU}(1,1)$, all of length $\frac{3\pi}{2}$. The figure shows geodesics starting at $\tilde{1}$, which, when parametrized by arc length, have initial values of α_a given (from left to right) by $-1, 0$ and $\frac{3}{4}$.

The vertical space (or vertical line, since it is one dimensional) in $\mathrm{SU}(1,1)$ with respect to 1 is

$$e^{\mathrm{span}\{Z\}} = \left\{ \left(\begin{array}{cc} e^{-i\vartheta} & 0 \\ 0 & e^{i\vartheta} \end{array} \right) \middle| \vartheta \in \mathbb{R} \right\} \cong U(1),$$

and the horizontal space is

$$e^{\mathrm{span}\{X,Y\}} = \{g = (z_1, z_2) \in \mathrm{SU}(1,1) | z_1 > 1\}.$$

Similarly, in $\mathrm{CSU}(1,1)$ the vertical line is defined by $w = 0$ (and from Proposition 1 it follows that K is isomorphic to $\mathbb{R}$), and the horizontal space is defined by $c = 0$.

4.2. **Length and number of geodesics.** Recall that χ_k is the unique number satisfying the equality $\tan \chi_k = \chi_k$ in the interval $\chi_k \in (\pi k, \pi k + \frac{\pi}{2})$.

Proposition 2. *Up to reparameterization, we have the following results regarding the number of geodesics connecting* $\tilde{1}$ *and* $\tilde{g} \in \mathrm{CSU}(1,1)$.

(a) *If* $\tilde{g}$ *is in the vertical line, that is* $\tilde{g} = (c, 0)$ *for some* c*, then the number of geodesics is uncountable (countably many geometrically different).*
(b) *If* $\tilde{g}$ *is in the horizontal space, then there exists a unique geodesic connecting* $\tilde{g}$ *and* $\tilde{1}$*. This geodesic is contained in the horizontal space.*
(c) *If* $\tilde{g}$ *is any other point, then we can obtain the number of geodesics in the following way. Let* k *be the largest positive integer, such that*

$$|c| \geq \sqrt{|w|^2 + \chi_k^2 + |w|^2\chi_k^2} - \chi_k - \tan^{-1}\left(\frac{\sqrt{|w|^2 + \chi_k^2 + |w|^2\chi_k^2} - \chi_k}{1 + \chi_k\sqrt{|w|^2 + \chi_k^2 + |w|^2\chi_k^2}} \right).$$

If there exists k, such that the above inequality is strict, then there are $2k+1$ geodesics connecting these two points. If k gives the equality, then the number of geodesics is $2k$. If no such k exists, then there is a unique geodesic.

Remark 4. *The number of geodesics in the case (c) above may be difficult to determine, but due to the fact that the value of* $\sqrt{|w|^2+\chi_k^2+|w|^2\chi_k^2}-\chi_k-\tan^{-1}\left(\frac{\sqrt{|w|^2+\chi_k^2+|w|^2\chi_k^2}-\chi_k}{1+\chi_k\sqrt{|w|^2+\chi_k^2+|w|^2\chi_k^2}}\right)$, *belongs to the interval* $\left(\pi k(\sqrt{|w|^2+1}-1),\ \pi(k+\frac{1}{2})(\sqrt{|w|^2+1}-1)\right)$, *this number remains between* $2k_1-1$ *an* $2k_1+1$, *where*

$$k_1=\left\lceil\frac{|c|}{\pi(\sqrt{1+|w|^2}-1)}-\frac{1}{2}\right\rceil.$$

The proof is a long case by case analysis, and therefore, we leave it to section 6. Regarding to (a), we say that two geodesics are *geometrically similar*, if one is the image of another under an isometry. The isometry considered in (a) is $(c,w)\mapsto(c,e^{i\theta}w)$, $\theta\in\mathbb{R}$.

Since $(\mathrm{CSU}(1,1),\tilde{D},\tilde{\rho}|_{\tilde{D}})$ is a step 2 regular sub-Riemannian manifold, there are no abnormal length minimizers. Every geodesic may be extended indefinitely, therefore the Carnot-Carathéodory metric is complete (see [21] and [22]). Hence, every point is connected to $\tilde{1}$ by a length minimizing geodesic. The same holds for $\mathrm{SU}(1,1)$. The above information along with the proof of Proposition 2, leads to the following result.

Corollary 1. (a) *If $\tilde{g}\in e^{\mathrm{span}\{Z\}}$, then $d(\tilde{1};\tilde{g})=|c|$.*

(b) *If $\tilde{g}\in e^{\mathrm{span}\{X,Y\}}$, then $d(\tilde{1};\tilde{g})=\sinh^{-1}|w|$.*

(c) *If $\tilde{g}=(c,w)$ is neither in the vertical, nor in the horizontal space, and $|c|<|w|-\tan^{-1}|w|$, then*

$$d(\tilde{1};\tilde{g})=\frac{1}{\beta}\left(|c|+\sin^{-1}\left(\frac{\beta|w|}{\sqrt{1+|w|^2}}\right)\right),$$

where $0<\beta<1$ is a unique number satisfying

$$\sqrt{\beta^{-2}-1}\left(|c|+\sin^{-1}\left(\frac{|w|\beta}{\sqrt{1+|w|^2}}\right)\right)=\sinh^{-1}\left(|w|\sqrt{1-\beta^2}\right).$$

(d) *If $\tilde{g}=(\pm(|w|-\tan^{-1}|w|),w)$, then $d(\tilde{1};\tilde{g})=|w|$.*

(e) *If $\tilde{g}=(c,w)$ satisfies the inequality $|w|-\tan^{-1}|w|<|c|\le\frac{\pi}{2}(\sqrt{|w|^2+1}-1)$, then*

$$d(\tilde{1};\tilde{g})=\frac{1}{\beta}\left(|c|+\sin^{-1}\left(\frac{|w|\beta}{\sqrt{1+|w|^2}}\right)\right),$$

where $1<\beta\le\frac{\sqrt{|w|^2+1}}{|w|}$ is a unique number satisfying

$$\sqrt{1-\beta^{-2}}\left(|c|+\sin^{-1}\left(\frac{|w|\beta}{\sqrt{1+|w|^2}}\right)\right)=\sin^{-1}\left(|w|\sqrt{\beta^2-1}\right).$$

(f) *If* $\tilde{g} = (c, w)$ *satisfies the inequality* $\frac{\pi}{2}(\sqrt{|w|^2+1} - 1) < |c|$*, then*

$$d(\tilde{1}; \tilde{g}) = \frac{1}{\beta}\left(|c| + \pi - \sin^{-1}\left(\frac{|w|\beta}{\sqrt{1+|w|^2}}\right)\right),$$

where $1 < \beta \leq \frac{\sqrt{|w|^2+1}}{|w|}$ *is a unique number satisfying*

$$\sqrt{1-\beta^{-2}}\left(|c| + \pi - \sin^{-1}\left(\frac{|w|\beta}{\sqrt{1+|w|^2}}\right)\right) = \pi - \sin^{-1}\left(|w|\sqrt{\beta^2 - 1}\right).$$

The details are again left to section 6. Notice that in all cases, the distance is independent of the sign of c, and of the argument of w. From these results for the universal cover, we easily obtain some conclusions for $\mathrm{SU}(1,1)$.

Corollary 2. *The number of geodesics connecting 1 and* $g \in \mathrm{SU}(1,1)$ *is*

(a) *uncountable (there are countably many geometrically different geodesics), if* $g \in e^{\mathrm{span}\{Z\}}$.
(b) *countable otherwise.*

The next result we prove here.

Corollary 3. *If* $g = (z_1, z_2) \in \mathrm{SU}(1,1)$*, then* $d(1; (z_1, z_2)) = d(\tilde{1}; (\operatorname{Arg} z_1, z_2))$.

Proof. Since both $\mathrm{SU}(1,1)$ and its cover are step 2 regular, it follows that $d(1; g) = \min_{\tilde{g}\in\pi^{-1}(g)}\{d(\tilde{1}; \tilde{g})\}$.

Let $|\operatorname{Arg} z_1| = c_0$ and let $c_k = c_0 + 2\pi k$, $k \in \mathbb{N}$. We need to show that $d(\tilde{1}; (c_j, w)) < d(\tilde{1}; (c_k, w))$, for $j < k$. When $|w| = 0$, this is a trivial consequence of Corollary 1, case (a).

If $|w| \neq 0$, we start with the case (f) $c_j > \pi(\sqrt{|w|^2 - 1} - 1)$. Then

$$d(\tilde{1}; (c_j, w)) \leq \sqrt{(c_j + k\pi - \tan^{-1}|w|)^2 - \pi^2} < \sqrt{(c_k + \pi - \frac{\pi}{2})^2 - \left(\frac{3\pi}{2}\right)^2} \leq d(\tilde{1}; (c_k, w)).$$

Let us consider the case (e) with $|w| - \tan^{-1}|w| < c_j < \frac{\pi}{2}(\sqrt{|w|^2 - 1} - 1)$. The value $d(\tilde{1}; (c_j, w)) \leq c_j + \frac{\pi}{2}$ is less than $\sqrt{(c_k + \tan^{-1}|w|)^2 - \frac{\pi^2}{4}}$ or $\sqrt{(c_k + \pi - \frac{\pi}{2})^2 - \left(\frac{3\pi}{2}\right)^2}$ (the lower bounds for $d(\tilde{1}; (c_k, w))$ are defined as follows: the first for $\pi(\sqrt{|w|^2 - 1} - 1) > c_k$, the second for $\pi(\sqrt{|w|^2 - 1} - 1) \leq c_k$)).

If $c_j = |w| - \tan^{-1}|w|$, then

$$d(\tilde{1}; (c_j, w)) = |w| = c_j + \tan^{-1}|w| < \sqrt{(|c_k| + \tan^{-1}|w|)^2 - \frac{\pi^2}{4}} \leq d(\tilde{1}; (c_k, w)).$$

If $c_j < |w| - \tan^{-1}|w|$, then we have the inequality

$$\sinh^{-1}|w| < d(1; (c_j, w)) < |w|,$$

which follows from the fact that $\frac{\sinh^{-1}(|w|\sqrt{1-\beta^2})}{\sqrt{1-\beta^2}}$ is increasing. So we only need to look at the case when $c_k < |w| - \tan^{-1}|w|$. We define β_j to be a number such that

$$g(\tilde{1},(c_j,w)) = \frac{1}{\beta_j}\left(c_j + \tan\left(\frac{|w|\beta}{\sqrt{1+|w|^2-|w|^2\beta^2}}\right)\right) = \frac{\sinh^{-1}(|w|\sqrt{1-\beta_j^2})}{\sqrt{1-\beta_j^2}}.$$

We argue from the contrary. Assume that $d(\tilde{1};(c_j,w)) \geq d(\tilde{1};(c_k,w))$. Then from the fact that $\frac{\sinh^{-1}(|w|\sqrt{1-\beta^2})}{\sqrt{1-\beta^2}}$ is increasing, it turns out that $\beta_j \geq \beta_k$. But then

$$d(\tilde{1};(c_j,w)) < \frac{1}{\beta_j}(c_j + \tan^{-1}|w|) < \frac{1}{\beta_k}(c_j + \tan^{-1}|w|) < \frac{c_k}{\beta_k} < d(\tilde{1};(c_k,w)),$$

which is a contradiction. □

4.3. **The cut and conjugate loci.** For sub-Riemannian Lie groups satisfying Theorem 3, we define a sub-Riemannian analogue of the exponential map about the identity 1 by

$$\exp_{sr} : \mathfrak{g} \setminus \mathfrak{k} \to G, \qquad X \mapsto e^X e^{-\operatorname{pr}_{\mathfrak{k}} X}.$$

For a more general definition of the exponential map in the sub-Riemannian setting, see [1]. We define *the conjugate locus* of G from the identity 1, as the set of critical values of $\exp_{sr}$. We often split the conjugate locus in several sets, defining the n-th conjugate locus by the set $\{\exp tX\}$, where $X \in \mathfrak{g} \setminus \mathfrak{k}$, and t is so that there exist exactly n values $0 < t_1 < \cdots < t_n = t$, so that $t_j X$ are all critical points. We define *the cut locus* from 1 as the set reachable by more than one minimizing geodesic.

Corollary 4. *The cut locus for* $\mathrm{CSU}(1,1)$ *is the vertical line. The cut locus of* $\mathrm{SU}(1,1)$ *consists of the points where either* $z_1 < 0$*, or* $z_2 = 0$*, and the points are different from the identity.*

Proof. The cut locus for $\mathrm{CSU}(1,1)$ follows from the proof of Proposition 2 and Corollary 1. For $\mathrm{SU}(1,1)$, if a point (z_1,z_2) is in the cut locus, then either $c = \operatorname{Arg} z_1, w = z_2$ is in the cut locus for $\mathrm{CSU}(1,1)$ (which is the set of points $(c,0)$ for arbitrary $c \neq 0$) or there exist more than one $\tilde{g} \in \mathrm{CSU}(1,1)$, such that $\pi(\tilde{g}) = g$ and $d(\tilde{1};\tilde{g}) = d(1;g)$. This only happens when $\operatorname{Arg} z_1 = \pi$. In this case there are exactly two points $\tilde{g}$. □

The following proposition was proved for $\mathrm{SU}(1,1)$ in [4] (for the isometric case of $\mathrm{SL}(2)$), but here we generalize it, including the universal cover $\mathrm{CSU}(1,1)$.

Proposition 3. *The n-th conjugate locus of* $\tilde{1} \in \mathrm{CSU}(1,1)$ *consists of the vertical line, if n is odd. If $n = 2j$, then it consists of the points given by the equation*

$$|c| = \sqrt{|w|^2+\chi_j^2+|w|^2\chi_j^2} - \chi_j - \tan^{-1}\left(\frac{\sqrt{|w|^2+\chi_j^2+|w|^2\chi_j^2}-\chi_j}{1+\chi_j\sqrt{|w|^2+\chi_j^2+|w|^2\chi_j^2}}\right).$$

Proof. First, observe that $d\exp_{sr}$ exists only for $\alpha_a \neq 0$. Put $w = u + iv$. Then we have

$$(c, u+iv) = \exp_{sr}(r\cos\theta X + r\sin\theta Y + a_3 Z)$$
$$= \left(-\phi(a) + a_3, -rS(r^2 - a_3^2)\cos(\theta + a_3) + irS(r^2 - a_3^2)\sin(\theta + a_3)\right).$$

The values of the elements of the matrix

$$\begin{pmatrix} \frac{\partial}{\partial\theta}c & ; & \frac{\partial}{\partial\theta}u & ; & \frac{\partial}{\partial\theta}v \\ \frac{\partial}{\partial r}c & ; & \frac{\partial}{\partial r}u & ; & \frac{\partial}{\partial r}v \\ \frac{\partial}{\partial a_3}c & ; & \frac{\partial}{\partial a_3}u & ; & \frac{\partial}{\partial a_3}v \end{pmatrix},$$

become

$$\begin{pmatrix} 0 & ; & rS\sin(\theta + a_3) & ; & rS\cos(\theta + a_3) \\ -\frac{a_3 r}{r^2 - a_3^2}\frac{1 - CS}{C^2 + a_3 S^2} & ; & -\frac{(r^2C - a_3^2 S)\cos(\theta + a_3)}{r^2 - a_3^2} & ; & \frac{(r^2C - a_3^2 S)\sin(\theta + a_3)}{r^2 - a_3^2} \\ 1 + \frac{a_3^2 - r^2CS}{(r^2 - a_3^2)(C^2 + a_3^2 S^2)} & ; & rS\sin(\theta) + \frac{ra_3(C - S)\cos(\theta + a_3)}{r^2 - a_3^2} & ; & rS\cos(\theta) - \frac{ra_3(C - S)\sin(\theta + a_3)}{r^2 - a_3^2} \end{pmatrix}.$$

Here, we have simplified by writing just S and C for $S(r^2 - a_3)$ and $C(r^2 - a_3^2)$. The determinant of the above matrix is $\frac{r^3 S(C-S)}{r^2 - a_3^2}$. The value of S vanishes at the points, for which $a_3^2 - r^2 = \pi^2 k^2, k \in \mathbb{N}$, and the image of such points is the vertical line (see the proof for Proposition 2 for more details). Moreover, $C - S$ vanishes only for $-\alpha = \chi_j^2$. Hence, for a generic $X \in \mathfrak{g} \setminus \mathfrak{k}$, the point tX can be singular for $\exp_{sr}$, only if $\alpha_a < 0$. Let us use the normalization $\alpha_a = -\pi^2$. Let the n-th value t_n be such that $t_n X$, is a singular point. Then it is clear that

$$t_{2j-1} = j, \qquad t_{2j} = \frac{\chi_j}{\pi}, \qquad j \in \mathbb{N}.$$

$\square$

Remark 5. *Let $(c, w) \in \mathrm{CSU}(1,1)$ be a point that does not belong to the vertical line (i.e. $w \neq 0$). Notice that if the value of $|c|$ from Proposition 2 increases (or the value of $|w|$ decreases), then (and only then) the number of geodesics increases when we pass through the even-indexed conjugate loci.*

Corollary 5. *The n-th conjugate locus of $1 \in \mathrm{SU}(1,1)$ for odd n is the vertical line, and if $n = 2j$, then it consist of the points*

$$z_1 = \frac{(-1)^j \left(1 \mp i\sqrt{|z_2|^2(1 + \chi_j^2) + \chi_j^2}\right) e^{\pm i\sqrt{|z_2|^2(1+\chi_j^2)+\chi_j^2}}}{\sqrt{1 + \chi_j^2}}.$$

5. Sub-Lorentzian geometry on CSU(1,1)

5.1. **Sub-Lorentzian maximizers and geodesics on** $\mathrm{CSU}(1,1)$**.** In contrast with the sub-Riemannian case, we only know that the relative maximizers exist locally. We have no guarantee of global existence of maximizers. Let us consider the distribution $\widetilde{E} =$

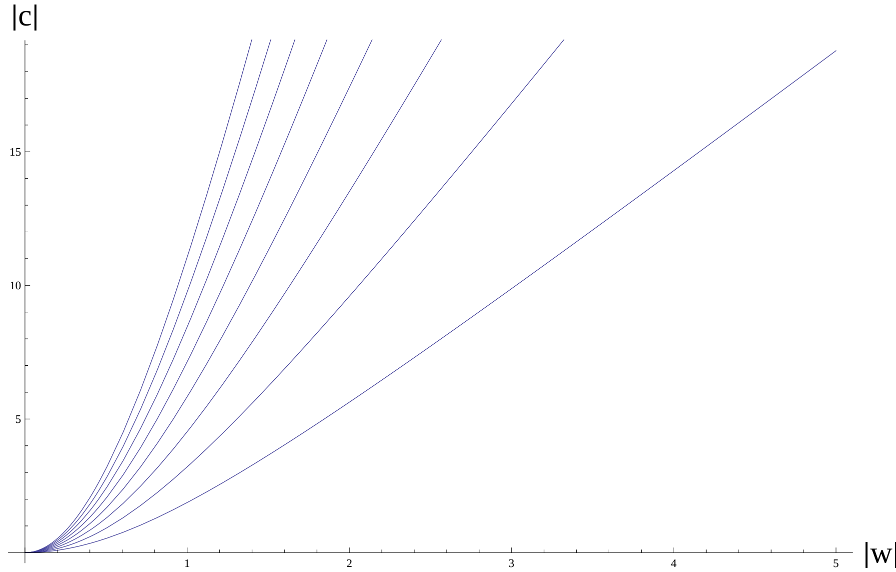

FIGURE 2. The above figure shows the eight first even conjugate loci. The horizontal axis corresponds to the values of $|w|$, and the values of $|c|$ occupy the vertical axis. Observe that a point under all the graphs is reachable by a unique geodesic. Then there are two geodesics in the 2-nd conjugate locus (that is the first graph). Between the 2-nd and the 4-th conjugate locus there are three, and so on.

$\operatorname{span}\{\widetilde{Y}, \widetilde{Z}\}$, with the metric $\tilde{\rho}$ restricted to $\widetilde{E}$. We formulate an optimal control problem of maximizing

$$J(u, \tilde{g}_u) = \int_0^\tau \sqrt{u_1^2 - u_2^2}\, dt, \qquad u \in U = \left\{ u \in \mathbb{R}^2 |\, u_1 = \sqrt{1 + u_2^2} \right\},$$

where $\dot{\tilde{g}}_u = f(u, \tilde{g}) = u_1 \widetilde{Z}(\tilde{g}) + u_2 \widetilde{Y}(\tilde{g})$.

We only do the case when $\varkappa = 1$. If $\varkappa = 0$ one gets the same results, so it follows that there are no strictly abnormal geodesics. In order to use PMP, let us define the pseudo-Hamiltonian

$$\mathscr{H}(u, p_1 X^* + p_2 Y^* + p_3 Z^*) = \sqrt{u_1^2 - u_2^2} + u_1 p_3 + u_2 p_2$$

$$= 1 + p_3 \sqrt{1 + u_2^2} + p_2 u_2.$$

where $u \in U = \{(u_1, u_2) \in \mathbb{R}^2 |\, u_1 = \sqrt{1 + u_2^2}\}$. For the existence of $H(p) = \max_{u \in U} \mathscr{H}(p, u)$, we need that $|p_3| > |p_2|$ and that $p_3 < 0$. In this case

$$H(p_1 X^* + p_2 Y^* + p_3 Z^*) = 1 - \sqrt{p_3^2 - p_2^2}.$$

This is from the fact that the optimal control is

$$u^* = \left(\frac{-p_3}{\sqrt{p_3^2 - p_2^2}}, \frac{p_2}{\sqrt{p_3^2 - p_2^2}} \right).$$

Since H cannot be extended to $TM \setminus s_0(M)$ (the tangent bundle with the zero-section removed), we solve the equation using the pseudo-Hamiltonian function $\mathscr{H}$ instead,

$$\begin{aligned} \dot{p}_1 &= \{p_1, \mathscr{H}(u,\lambda)\}|_{u=u^*} = -2(u_1 p_2 + u_2 p_3)|_{u=u^*} = 0, \\ \dot{p}_2 &= \{p_2, \mathscr{H}(u,\lambda)\}|_{u=u^*} = 2u_1 p_1|_{u=u^*} = -\frac{2p_1p_3}{\sqrt{p_3^2 - p_2^2}}, \\ \dot{p}_3 &= \{p_3, \mathscr{H}(u,\lambda)\}|_{u=u^*} = -2u_2 p_1|_{u=u^*} = -\frac{2p_1p_2}{\sqrt{p_3^2 - p_2^2}}, \end{aligned}$$

$p_3^2 - p_2^2$ is a first integral in this case, and from the condition $\mathscr{H}(u, p_1X^* + p_2Y^* + p_3Z_3) = 0$, $p_3^2 - p_2^2$ is equal to 1. We find the solutions (denote $p_{jo} := p_j(0)$)

$$\begin{aligned} p_1(t) &\equiv p_{1o}, \\ p_2(t) &= p_{2o} \cosh(2p_{1o}t) - p_{3o} \sinh(2p_{1o}t), \\ p_3(t) &= p_{3o} \cosh(2p_{1o}t) - p_{2o} \sinh(2p_{1o}t). \end{aligned}$$

Observe that $|p_{3o}| > |p_{2o}|$, and $p_{3o} < 0$. Note also that p_{3o} has to be equal to $-\sqrt{1+p_{2o}^2}$. In order to simplify our calculation, we first solve it for $\mathrm{SU}(1,1)$, and then lift it to $\mathrm{CSU}(1,1)$. We have to solve the following differential equation.

$$\dot{g} = \left.\frac{\partial \mathscr{H}}{\partial p}\right|_{u=u^*} = -p_3 Z(g) + p_2 Y(g), \tag{1}$$

Make the following observation that

$$e^{a_1 X} = \begin{pmatrix} \cosh a_1 t & -\sinh a_1 t \\ -\sinh a_1 t & \cosh a_1 t \end{pmatrix}.$$

From this we notice that

$$\begin{aligned} e^{-a_1X} Z = Z e^{a_1 X} &= \cosh a_1 Z + \sinh a_1 Y, \\ e^{-a_1X} Y = Y e^{a_1 X} &= \cosh a_1 Y + \sinh a_1 Z, \\ e^{a_1X} X = X e^{a_1 X} &= \cosh a_1 X + \sinh a_1 \cdot 1. \end{aligned}$$

If we expand equation (1), we can write

$$\begin{aligned} \dot{g} &= -\left(p_{3o}\cosh(2p_{1o}t) - p_{2o}\sinh(2p_{1o}t)\right) Z(g) + \left(p_{2o}\cosh(2p_{1o}t) - p_{3o}\sinh(2p_{1o}t)\right) Y(g) \\ &= g(-p_{3o} Z e^{2tp_{1o}X} + p_{2o} Y e^{2tp_{1o}X}) = g e^{-tp_1 X}(-p_{3o}Z + p_{2o}Y) e^{tp_{1o}X}, \end{aligned}$$

and multiplying from the right by e^{-ta_1X} and adding $-p_{1o} g X e^{-tp_{1o}X} = -p_{1o} g e^{-tp_{1o}X} X$ on both sides, we obtain

$$\dot{g} e^{-tp_{1o}X} - p_{1o} g X e^{-tp_{1o}X} = \frac{\partial}{\partial t}\left(g e^{-tp_{1o}X}\right) = g e^{-tp_{1o}X}(-p_{3o}Z + p_{2o}Y - p_{1o}X).$$

It follows that

$$ge^{-tp_{1o}X} = e^{t(-p_{1o}X+p_{2o}Y-p_{3o}Z)},$$

i.e.,

$$g = e^{t(-p_{1o}X+p_{2o}Y-p_{3o}Z)}e^{tp_{1o}X}.$$

If we lift this curve and consider Theorem 3, then we easily get to the following proposition.

Proposition 4. *Assume that there exists a global length maximizing curve between $\tilde{1}$ and $\tilde{g}$. Then, this curve is a timelike, future-directed normal geodesic on the form*

$$e^{t(a_1X+a_2Y+a_3Z}e^{-ta_1X},$$

where $a_3 > 0$ and $|a_2| < a_3$.

Explicitly, these geodesics have the form

$$\begin{aligned} c(t) = & - \phi(ta) \\ & + \tan^{-1}\left(\frac{a_2C(\alpha_a,t)S(\alpha_a,t) - a_1a_3S^2(\alpha_a,t)}{(1+|a_1+ia_2|^2|S(\alpha_a,t)|^2)\coth(a_1t) - a_1C(\alpha_a,t)S(\alpha_a,t) - a_2a_3S^2(\alpha_a,t)}\right), \\ w(t) &= (C(\alpha_a,t) - ia_3S(\alpha_a,t))\sinh(a_1t) - (a_1 - ia_2)S(\alpha_a,t)\cosh(a_1t). \end{aligned}$$

The projection of them to $\mathrm{SU}(1,1)$ is given by

$$z_1 = (C(\alpha_a,t) - ia_3S(\alpha_a,t))\cosh(a_1t) - (a_1 - ia_2)S(\alpha_a,t)\sinh(a_1t),$$
$$z_2 = (C(\alpha_a,t) - ia_3S(\alpha_a,t))\sinh(a_1t) - (a_1 - ia_2)S(\alpha_a,t)\cosh(a_1t).$$

We will discuss the situation when these curves correspond to length maximizers in subsection 5.3.

5.2. **Number of geodesics.** The result in the sub-Lorentzian case is more complicated, than in the sub-Riemannian case. So we have to give some definitions in order to describe geodesics in a reasonable way. First, let us define a function $f_\pm : \mathbb{R}\times\mathbb{N}_0 \to \mathbb{R}$, by

$$f_\pm(s;k) = \begin{cases} \sqrt{1-s^2} & \text{if } |s|<1, \\ \mp\sinh\left(\frac{(2k\pm1)\pi}{2}\sqrt{s^2-1}\right) & \text{if } |s|\geq 1. \end{cases}$$

Also, we define the numbers $\omega_k, k\in\mathbb{N}$ as the numbers satisfying the equations

$$\frac{\omega_k}{\sqrt{1-\omega_k^2}} - \sin^{-1}\omega_k = \pi k, \qquad 0<\omega_k<1.$$

Finally, we define the function

$$F(s,\omega) = \sqrt{1-\omega^2}\cosh\sqrt{\frac{s^2-\omega^2}{1-\omega^2}} - \sqrt{s^2-\omega^2}\sinh\sqrt{\frac{s^2-\omega^2}{1-\omega^2}}.$$

Let us construct the following subsets of $\mathrm{CSU}(1,1)$: Ω_0 consist of all points

$$c = \mathrm{Arg}\left(x_1 - i\sqrt{s^2+y_2^2}\right), \qquad w = \pm\sqrt{x_1^2+s^2-1}+iy_2,$$

where $y_2 \in \mathbb{R}$, $s > 0$, and $x_1 \in (-\infty, -\sqrt{1-s^2}] \bigcup [\sqrt{1-s^2}, 1 - \frac{s^2}{2})$ when $s < 1$, and $x_1 \in (-\infty, 1 - \frac{s^2}{2})$ otherwise. Further, for $k \in \mathbb{N}$, define Ω_k as a set of all points

$$c = -\pi k + \operatorname{Arg}\left(x_1 - i\sqrt{s^2 + y_2^2}\right), \qquad w = \pm\sqrt{x_1^2 + s^2 - 1} + iy_2,$$

where $y_2 \in \mathbb{R}$, $s > 0$, and $x_1 \in \mathbb{R} \setminus \left(-\sqrt{1-s^2}, \sqrt{1-s^2}\right)$ for $s < 1$, and $x_1 \in (-\infty, F(s, \omega_k)]$ otherwise. We split them into the following sets

- A_k consist of all points
$$c = -\pi k + \operatorname{Arg}\left(x_1 - i\sqrt{s^2 + y_2^2}\right), \qquad w = \pm\sqrt{x_1^2 + s^2 - 1} + iy_2,$$
  where $y_2 \in \mathbb{R}$, $s > 0$, and
  - $x_1 \in \mathbb{R} \setminus \left(-\sqrt{1-s^2}, \sqrt{1-s^2}\right)$, when $s \leq \omega_k$,
  - $x_1 \in (-\infty, -\sqrt{1-s^2}] \cup (F(s, \omega_k), \infty)$, when $\omega_k < s < 1$,
  - $x_1 \in (-\infty, 0] \cup \{F(1, \omega_k)\}$, when $s = 1$,
  - $x_1 \in \{F(s, \omega_k)\}$, when $s > 1$.
- B_k consist of all points
$$c = -\pi k + \operatorname{Arg}\left(x_1 - i\sqrt{s^2 + y_2^2}\right), \qquad w = \pm\sqrt{x_1^2 + s^2 - 1} + iy_2,$$
  where $y_2 \in \mathbb{R}$, $s > 0$, and
  - $x_1 \in \{\sqrt{1-s^2}, F(s, \omega_k)\}$, when $s < \omega_k$,
  - $x_1 \in (0, F(s, \omega_k))$, when $s = 1$,
  - $x_1 \in (-\infty, F(s; k))$, when $s > 1$.
- C_k consist of all points
$$c = -\pi k + \operatorname{Arg}\left(x_1 - i\sqrt{s^2 + y_2^2}\right), \qquad w = \pm\sqrt{x_1^2 + s^2 - 1} + iy_2,$$
  where $y_2 \in \mathbb{R}$, $\omega_k < s < 1$, and $x_2 \in (\sqrt{1-s^2}, F(s, \omega_k))$.

Let $\Xi_k = \{(-\pi k, x_2) \in \mathrm{CSU}(1,1) |\, x_2 \in \mathbb{R}\}$. Finally, define

$$\Omega := \Omega_0 \cup \bigcup_{k=1}^{\infty} (\Omega_k \cup \Xi_k), \qquad \Xi := \bigcup_{k=1}^{\infty} \Xi_k,$$

$$A := \Omega_0 \cup \bigcup_{k=1}^{\infty} A_k, \qquad B := \bigcup_{k=1}^{\infty} B_k, \qquad C := \bigcup_{k=1}^{\infty} C_k.$$

Proposition 5. *There are timelike, future directed geodesics connecting $\tilde{1}$ and $\tilde{g}$, only if $\tilde{g} \in \Omega$. More precisely,*

- *the geodesic is unique if $\tilde{g} \in A$;*
- *there are two geodesics if $\tilde{g} \in B$;*
- *there are three geodesic if $\tilde{g} \in C$;*
- *there are countable many geodesics if $\tilde{g} \in \Xi$.*

This case by case analysis is left to section 6.

Remark 6. *Let χ_k be the constants defined in subsection 3.1. Then $\omega_k = |\sin\chi_k|$, so these numbers play a role here as well. However, unlike what was the case when we considered the sub-Riemannian geodesics, the mapping*

$$a_1X + a_2Y + a_3Z \mapsto e^{a_1X+a_2Y+a_3Z}e^{-a_1X},$$

does not have any critical points satisfying $\alpha_a = -\chi_k^2$.

We also remark the following interesting comparison with Lorentzian geometry.

Proposition 6. *The set reachable by Lorentzian geodesics starting at $\tilde{1}$, is the planes $\{(c,w) \in \mathrm{CSU}(1,1) \,|\, c = -\pi k - \frac{\pi}{2}, k \in \mathbb{N}\}$, along with all points satisfying $c < 0$, $|w| < \tan c$, and $(c,w) = (-\pi k, 0), k \in \mathbb{N}$. More precisely,*

(a) *there are uncountably many geodesics if $w = 0$ (all geometrically similar, except for the case $c = 0 \mod \pi$, when they are uncountably many geometrically different).*
(b) *there is a unique geodesic if $w \neq 0$.*

Proof. A time-like geodesic $\tilde{g}(t)$, starting at $\tilde{1}$ satisfies $\tilde{g}(1) = e^{a_1X+a_2Y+a_3Z}$, and we know that $a_3 > 0$ and $a_3^2 > a_1^2 + a_2^2$ (and hence $\alpha_a < 0$). We rewrite the geodesics in terms of the parameters k, α', r and θ, where $\sqrt{-\alpha_a} = \pi k + \alpha', k \in \mathbb{N}, \alpha' \in (-\frac{\pi}{2}, \frac{\pi}{2}]$ and $a_1 + ia_2 = re^{i\theta}$. The geodesics in these coordinates look like

$$c = -\tan^{-1}\left(\frac{\sqrt{r^2 + (\pi k + \alpha')^2}}{\pi k + \alpha'}\tan\alpha'\right) - k\pi, \qquad w = -\frac{r}{\pi k + \alpha'}e^{i\theta}(-1)^k\sin\alpha'.$$

Now θ can be determined by $\arg w$, and $k = \lceil -\frac{c}{\pi} - \frac{1}{2}\rceil$. Rewriting the equation for $|w|$ as

$$|w|^2 + \sin^2\alpha' = \frac{r^2 + (\pi k + \alpha')^2}{(\pi k + \alpha')^2}\sin^2\alpha',$$

and inserting it in the equation for c, we get that

$$\alpha' = -\sin^{-1}\left(\tan(c)\sqrt{1 - \frac{|w|^2}{\tan^2 c}}\right),$$

for $c \neq \frac{\pi n}{2}$. For the remaining cases, $\alpha' = 0$ for $c = \pi n$, and $\alpha' = \frac{\pi}{2}$ for $c = \frac{\pi}{2} + \pi n$. It is obvious that there are solutions only if $|\tan c| \geq |w|$. The equality can never be attained, because $\operatorname{sgn}\,\tan c = -\operatorname{sgn}\,\alpha'$, which follows from

$$\tan c = -\frac{\sqrt{r^2 + (\pi k + \alpha')^2}}{\pi k + \alpha'}\tan\alpha'.$$

The equation

$$r^2 = \frac{(\pi k + \alpha')^2|w|}{\sin^2\alpha'},$$

determines r for $\alpha' \neq 0$. When $\sin\alpha' = 0$, there are no restrictions on r. Also, if $|w| = 0$, there are no restrictions on θ. $\square$

Proposition 6 yields that the set of points reachable by the Lorentzian geodesics starting at $\tilde{1}$ neither is contained, nor contains the set reachable from $\tilde{1}$ by the sub-Lorentzian geodesics. This contrasts the fact that the Lorentzian timelike future always contains the sub-Lorentzian one.

5.3. **Lorentzian and sub-Lorentzian timelike future.** Using the information we have collected, concerning sets reachable by geodesics, we obtain some results for the timelike future in general.

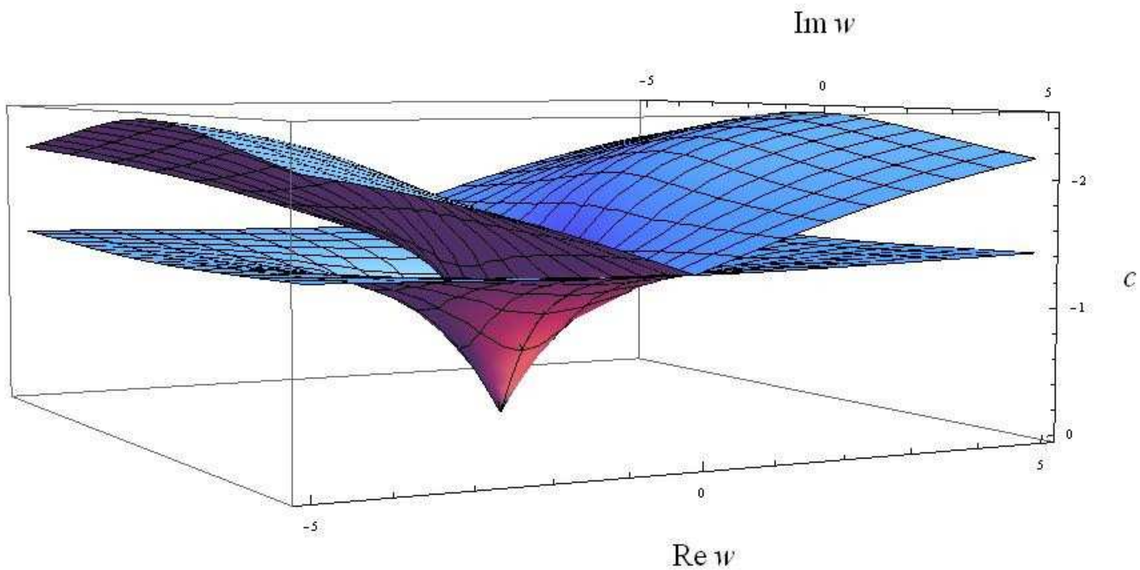

FIGURE 3. The above figure shows the lower bounds of Lorentzian and sub-Lorentzian timelike future for CSU(1, 1), showing the sub-Lorentzian future contained in the Lorentzian one. Remark that along the imaginary w-axis, $\mathcal{I}^+(\tilde{1}, \tilde{\rho}|_{\tilde{E}})$ is tangent to $\mathcal{I}^+(\tilde{1}, \rho)$, and the negative direction of c is set upward.

Proposition 7. (a) *The timelike future of* CSU(1, 1) *with respect to the Lorentzian metric* $\tilde{\rho}$, *is given by*

$$\mathcal{I}^+(\tilde{1}, \tilde{\rho}) = \left\{(c, w) \in \mathrm{CSU}(1, 1) |\, c < -\tan^{-1} |w|\right\}.$$

(b) *The timelike future of* CSU(1, 1) *with respect to the sub-Lorentzian metric* $\tilde{\rho}|_{\tilde{E}}$, *is given by*

$$\mathcal{I}^+(\tilde{1}, \tilde{\rho}|_{\tilde{E}}) = \{(c, w) \in \mathrm{CSU}(1, 1) |\, c < \mathrm{Arg}(1 - |\mathrm{Re}\ w| - i\sqrt{(\mathrm{Im}\ w)^2 + 2|\mathrm{Re}\ w|})\}.$$

Proof. Let us denote by $Q = \{(c, w) \in \mathrm{CSU}(1, 1) |\, c < -\tan^{-1} |w|\}$, and

$$\widehat{Q} = \{(c, w) \in \mathrm{CSU}(1, 1) |\, c < \mathrm{Arg}(1 - |\mathrm{Re}\ w| - i\sqrt{(\mathrm{Im}\ w)^2 + 2|\mathrm{Re}\ w|})\}.$$

In both sub-Lorentzian and Lorentzian settings, we show by using the information about the geodesics, that Q and $\widehat{Q}$ are included in their respective timelike futures. To show the

opposite inclusion, remark that in both Lorentzian and sub-Lorentzian settings, we use the fact that there exist relative length maximizers locally. So there exists a neighborhood V of $\tilde{1}$, such that

$$\mathcal{I}^+(\tilde{1}, \tilde{\rho}|_V) = V \cap Q.$$

Define $U = V \cap Q$. Now from left invariance, we know that $\mathcal{I}^+(\tilde{g}, \tilde{\rho}) = L_{\tilde{g}}\mathcal{I}^+(\tilde{1}, \tilde{\rho})$, and it follows that

$$\mathcal{I}^+(\tilde{1}, \tilde{\rho}) = \bigcup_{\tilde{g}\in\mathcal{I}^+(\tilde{1},\tilde{\rho})} L_{\tilde{g}}U.$$

Hence it is sufficient to show that $L_{\tilde{g}}U \subseteq Q$, for every $\tilde{g} \in Q$. Since every timelike curve from $\tilde{1}$ to a point outside Q has to pass through ∂Q, it follows by continuity that it is sufficient to show that $L_{\tilde{g}_0}\overline{U} \subseteq L_{g_0}\overline{Q} \subseteq \overline{Q}$, for every $\tilde{g}_0 \in \partial Q$. Here, $\overline{U}$ means the closure of U. Finally, since $L_{\tilde{g}_0}$ is an isomorphism, we have that $\partial L_{\tilde{g}_0}\overline{Q} = L_{\tilde{g}_0}\partial Q$, so all remaining arguments turn down to show that $\tilde{g}_0\tilde{g} \in \overline{Q}$ for every $\tilde{g}_0, \tilde{g} \in \partial Q$. The same holds for $\widehat{Q}$.

In order to prove (a) we show that there are timelike geodesics connecting every point satisfying $-\frac{\pi}{2} \leq c < -\tan|w|$ with the origin. Let (c, w) be an arbitrary point satisfying $c < -\frac{\pi}{2}$. Then we can construct a timelike curve from $\tilde{1}$, by taking a geodesic from $\tilde{1}$ to $(\frac{\pi}{2}, iwe^{-ic})$, and then continue by a left translation of a geodesic from $\tilde{1}$ to $(c + \frac{\pi}{2}, 0)$.

Let $\tilde{g}_0, \tilde{g} \in \partial Q$, $\tilde{g}_0 = (-\tan^{-1} r_0, r_0e^{i\theta_0}), \tilde{g} = (-\tan^{-1} r, re^{i\theta})$. First, if $r \geq 1$ and $r_0 \geq 1$, then

$$\left(-\tan^{-1} r_0, r_0e^{i\theta_0}, -\tan^{-1} r, re^{i\theta}\right) = (C, W),$$

$$C = -\tan^{-1} r_0 - \tan^{-1} r - \tan^{-1}\left(\frac{(rr_0 - 1)\sin\vartheta + (r + r_0)\cos\theta}{\frac{1}{rr_0} + \frac{r}{r_0} + \frac{r_0}{r} + rr_0 + (r + r_0)\sin\vartheta - (rr_0 - 1)\cos\vartheta}\right),$$

$C \leq -\frac{\pi}{2}$, so in this case $(C, W) \in Q$. Here $\vartheta = \theta_0 - \theta$.

We now turn to the case when either r_0 or r is less than 1. If we denote

$$(x_1 + iy_1, z_2) := \pi(\tilde{g}_0)\pi(\tilde{g}),$$

then we have

$$\begin{aligned}
x_1 &= 1 - rr_0 + rr_0\cos\vartheta,\\
y_1 &= -r_0 - r + rr_0\sin\vartheta,\\
|z_2|^2 &= r^2 + r_0^2 + 2r^2r_0^2 + 2rr_0\left((1 - rr_0)\cos\vartheta - (r + r_0)\sin\vartheta\right)\\
&= 2(1 - rr_0)x_1 + 2(r + r_0)|y_1| - (r + r_0)^2 - 2(1 - rr_0).
\end{aligned}$$

Since $r < 1$ or $r_0 < 1$, we conclude that $y_1 < 0$. It follows that $\operatorname{Arg}(x_1 + iy_1) \leq -\tan^{-1}|z_2| = \operatorname{Arg}(1 - i|z_2|)$, because $x_1 \leq 1$. As a consequence, $g_0g \in \overline{Q}$.

Turning to (b), let us observe that Ω_0 consist of all points, satisfying

$$-\pi + \tan^{-1}\left(\frac{(\operatorname{Im} w)^2}{1 + (\operatorname{Re} w)^2}\right) < c < \operatorname{Arg}(1 - |\operatorname{Re} w| - i\sqrt{(\operatorname{Im} w)^2 + 2|\operatorname{Re} w|}).$$

Clearly all the points in Ω_0 are in the timelike future. Pick up any point (c, w) satisfying $c < \operatorname{Arg}(1-|\operatorname{Re} w|-i\sqrt{(\operatorname{Im} w)^2 + 2|\operatorname{Re} w|})$. Let γ be a left translation by $(0, we^{-ic})$ of any geodesic connecting $\tilde{1}$ and $(c, 0)$. The endpoint of γ is (c, w). Since $c < \operatorname{Arg}(1 - |\operatorname{Re} w| -$

$i\sqrt{(\mathrm{Im}\ w)^2+2|\mathrm{Re}\ w|})$, the curve γ must at some point intersect Ω_0. Pick up any point $\tilde{g}_1 \in \gamma \cap \Omega_0$, travel from $\tilde{1}$ to $\tilde{g}_1$ along a geodesic, and continue along γ to (c,w).

Let $\tilde{g}_0, \tilde{g} \in \widehat{Q}$ be of the form

$$g_0 = \left(-\mathrm{Arg}(1-|x|-i\sqrt{y^2+2|x|}), x+iy\right), \quad \tilde{g} = \left(-\mathrm{Arg}(1-|u|-i\sqrt{v^2+2|u|}), u+iv\right).$$

Assume first that $|u| \leq \frac{1}{2}$. If we again denote $\pi(\tilde{g}_0)\pi(\tilde{g}) = (x_1+iy_1, x_2+iy_2)$, we then obtain

$$\begin{aligned} x_1 &= (1-|x|)(1-|u|) + xu + vy - \sqrt{y^2+2|x|}\sqrt{v^2+2|u|}, \\ x_2 &= (1-|x|)u + v\sqrt{y^2+2|x|} + x(1-|u|) - y\sqrt{v^2+2|u|}. \end{aligned}$$

We only need to show that $1-x_1-|x_2| \geq 0$. Denote $\sigma = \mathrm{sgn}\ x_2$. Then

$$\begin{aligned} 1-x_1-|x_2| &= (|x|-\sigma x)+(|u|-\sigma u)-(|u|-\sigma u)(|x|-\sigma x)+(\sqrt{v^2+2|u|}-\sigma v)(\sqrt{y^2+2|x|}+\sigma y) \\ &\geq (|x|-\sigma x)+(|u|-\sigma u)-(|u|-\sigma u)(|x|-\sigma x) \geq 0, \end{aligned}$$

since $0 \leq (|u|-\sigma u) \leq 1$. If $|u| > \frac{1}{2}$, the equality

$$\left(-1-|u|-i\sqrt{v^2+2|u|}, u+iv\right)$$

$$= \left(1-\frac{i}{2}\frac{1}{\sqrt{v^2+2|u|}+\sigma_1 v}, -\frac{i}{2}\frac{\sigma_1}{\sqrt{v^2+2|u|}+\sigma_1 v}\right)(1-|u_1|-i\sqrt{v_1^2+2|u_1|}, u_1+iv_1),$$

where $\sigma_1 = \mathrm{sgn}\ u$, and

$$u_1 = u - \frac{\sigma_1}{2}, \qquad v_1 = \frac{\sigma_1}{2}\left(\sqrt{v^2+2|u|}+\sigma_1 - \frac{2|u|-1}{\sqrt{v^2+2|u|}-\sigma_1 v}\right),$$

tells us that $\tilde{g}$ may be written as a finite product of boundary elements, all with real part of the second coordinate of absolute value less than or equal to $\frac{1}{2}$. $\square$

Since we know that locally, the Lorentzian or sub-Lorentzian causal future is the closure of the timelike future (see [13]), it follows from left invariance that $\mathcal{J}^+(\tilde{1},\tilde{\rho})$ and $\mathcal{J}^+(\tilde{1},\tilde{\rho}|_{\tilde{E}})$ are the closure of $\mathcal{I}^+(\tilde{1},\tilde{\rho})$ and $\mathcal{I}^+(\tilde{1},\tilde{\rho}|_{\tilde{E}})$, respectively.

Lemma 1. *Both with respect to the Lorentzian $\tilde{\rho}$ and the sub-Lorentzian metric $\tilde{\rho}|_{\tilde{E}}$,* $\mathrm{CSU}(1,1)$ *is strongly causal.*

Proof. The proof is the same for both the Lorentzian and sub-Lorentzian cases. By left invariance, it is sufficient to find a strongly causal neighborhood. Let $U_{c_0}, 0 < c_0 < \frac{\pi}{2}$ be the set of all points (c,w) in the timelike future of $\tilde{1}$ satisfying $c > -c_0$. Observe that inequality $|w| < \tan|c|$ must hold at the same time. Since, the causal future of all the elements in U_{c_0} is contained in the causal future of $\tilde{1}$, any causal curve must exit through the surface $\{(-c_0,w_0) \,|\, |w_0| \leq \tan^{-1} c_0\}$. Then we need to show that for any such element $L_{(c_0,w_0)}U_{c_0} \cap U_{c_0} = \emptyset$. Observe that if $(c,w) \in U_{c_0}$ and $(-c_0,w_0)(c,w) = (C,W)$, then

$$C = -c_0 + c + \mathrm{Arg}(\sqrt{(1+|w_0|^2)(1+|w|^2)} + w_0 w e^{-i(c+c_0)})$$

$$\leq -c_0 + c + \tan^{-1}\left(\frac{|w_0 w|}{\sqrt{(1+|w_0|^2)(1+|w|^2)}}\right) < -c_0 + c + \tan^{-1}(\sin c_0 \sin |c|)$$
$$< -c_0 + c + \tan^{-1}(\sin |c|) < c_0.$$

Hence, U_{c_0} is strongly causal neighborhood of its elements. □

Proposition 8. *The distance between $\tilde{1}$ and (c, w) with respect to the Lorentzian metric $\tilde{\rho}$ is equal to*

$$d(\tilde{1}; (c, w)) = \begin{cases} 0 & \text{if } c \geq -\tan^{-1}|w| \\ \sin^{-1}\left(\sqrt{\tan^2 c - |w|^2}\right) & \text{if } -\frac{\pi}{2} < c < -\tan^{-1}|w| \\ \frac{\pi}{2} & \text{if } c = -\frac{\pi}{2} \\ \pi - \sin^{-1}\left(\sqrt{\tan^2 c - |w|^2}\right) & \text{if } -\pi + \tan^{-1} < c < -\frac{\pi}{2} \end{cases}$$

When $c \leq -\pi + \tan^{-1}|w|$, we know that $d(\tilde{1}; (c, w)) \geq \pi$.

Proof. The identity for $c \geq -\tan^{-1}|w|$ is trivial. For $-\pi + \tan^{-1}|w| < c < -\tan^{-1}|w|$, the formula follows from the fact that this space is globally hyperbolic, and from the proof of Proposition 6. The lower bound for the distance when $c \leq -\pi + \tan^{-1}|w|$ follows from the reverse triangle inequality, and from the fact that every such element has points in its timelike past of distance arbitrarily close to π.

To prove that $-\pi + \tan^{-1}|w| < c < -\tan^{-1}|w|$, we need to show that for every pair $\tilde{g}_1, \tilde{g}_2$ in this subset, $\mathcal{J}^+(\tilde{g}_1, \tilde{\rho}) \cap \mathcal{J}^-(\tilde{g}_2, \tilde{\rho})$ is compact. It is trivial that $\mathcal{J}^+(\tilde{g}_1, \tilde{\rho}) \subset \mathcal{J}^+(\tilde{1}, \tilde{\rho})$, and by the same reasoning $\mathcal{J}^-(\tilde{g}_2, \tilde{\rho}) \subseteq \mathcal{J}^-((c_0, 0), \tilde{\rho})$. for some $-\pi < c_0 < 0$. Clearly,

$$\mathcal{J}^-((c_0, 0), \tilde{\rho}) = \{(c, w) \in \mathrm{CSU}(1, 1) |\, c \geq c_0 + \tan^{-1}|w|\}.$$

It follows that $\mathcal{J}^+(\tilde{1}, \tilde{\rho}) \cap \mathcal{J}^-((c_0, 0), \tilde{\rho})$, is compact for $-\pi < c_0 < 0$, but not when $c_0 \leq -\pi$. □

One would expect an analogous statement for the sub-Lorentzian metric presenting a description of the distance function in Ω_0, given the proof of Proposition 5, but it is more difficult to prove whether it is globally hyperbolic or not.

6. Proofs of main results

6.1. **Proof of Proposition 2.** The technique of this proof is to consider a general geodesic satisfying $\tilde{g}(0) = \tilde{1}$ and $\tilde{g}(1) = (c, w)$. Each geodesic is determined by their initial conditions (a_1, a_2, a_3), where $a_1^2 + a_2^2 \neq 0$. Our task will be, given the final point of the geodesic, (c, w), to find how many choices of initial conditions do we have.

It will be practical to use $re^{i\theta} := a_1 + ia_2$. Notice that $\alpha_a = r^2 - a_3^2$, and hence it is independent of θ. We need to solve the equations

$$c = -\phi(a) + a_3, \qquad |w| = r|S(r^2 - a_3^2)| \tag{2}$$

$$\frac{w}{|w|} = -e^{-i\theta}\frac{S(r^2 - a_3^2)}{|S(r^2 - a_3^2)|}, \qquad \text{if } w \neq 0. \tag{3}$$

We start by proving a). If $w = 0$, then $S(\alpha_a) = 0$ (and only then), so $\sqrt{a_3^2 - r^2} = \pi k$, where $k \in \mathbb{N}$. Furthermore, $\operatorname{sgn}(c) = \operatorname{sgn}(a_3)$, $|c| = |a_3| - \pi k$ and $r = \sqrt{a_3^2 - \pi^2 k^2} = \sqrt{c(c + 2\pi k)}$, so there are countably many choices. Also, there are no restrictions to θ.

To obtain the remaining results, let us define $\beta := \frac{a_3}{r}$. It is easy to see that (3) determines $e^{i\theta}$, so we only need to look at the number of solutions to (2). If $\beta = 0$ (i.e. $a_3 = 0$), then the geodesics are contained in the horizontal space, and there is a unique one for every point ($r = \sinh^{-1}|w|$).

If $\beta \neq 0$, then $\alpha_a = a_3^2(\beta^{-2} - 1)$, and this to leads to the remaining cases.

6.1.1. *The case* $\alpha_a > 0 \iff 0 < |\beta| < 1$. In order to find a solution (2), we must solve the equations

$$c = a_3 - \tan^{-1}\left(\frac{\beta}{\sqrt{1-\beta^2}} \tanh\left(\frac{a_3\sqrt{1-\beta^2}}{\beta}\right)\right), \tag{4}$$

$$\sinh\left(\frac{a_3\sqrt{1-\beta^2}}{\beta}\right) = |w|\sqrt{1-\beta^2}. \tag{5}$$

Observe that $c - \frac{\pi}{2} < a_3 < c + \frac{\pi}{2}$. We rewrite (5) as

$$\tanh\left(\frac{a_3\sqrt{1-\beta^2}}{\beta}\right) = \sqrt{\frac{|w|^2 - |w|^2\beta^2}{1 + |w|^2 - |w|^2\beta^2}},$$

and inserting it in (4), we obtain

$$a_3 = c + \tan^{-1}\left(\frac{\beta|w|}{\sqrt{1 + |w|^2 - |w|^2\beta^2}}\right).$$

Substituting a_3 back into (5), we have that β is a solution to the equation

$$B(\beta) \equiv \frac{\sqrt{1-\beta^2}}{\beta}\left(c + \tan^{-1}\left(\frac{\beta|w|}{\sqrt{1 + |w|^2 - |w|^2\beta^2}}\right)\right) - \sinh^{-1}(|w|\sqrt{1-\beta^2}) = 0, \tag{6}$$

which has a solution only for $0 < |c| < |w| - \tan^{-1}|w|$, and it is unique. To show this, take the derivative of $B(\beta)$ as

$$\frac{\partial B}{\partial \beta} = -\frac{1}{\beta^2\sqrt{1-\beta^2}}\left(c + \tan^{-1}\left(\frac{|w|\beta}{\sqrt{1 + |w|^2 - |w|^2\beta^2}}\right) - \frac{|w|\beta}{\sqrt{1 + |w|^2 - |w|^2\beta^2}}\right). \tag{7}$$

The limits

$$\lim_{|\beta| \to 1-} B(\beta) = 0, \quad \lim_{|\beta| \to 0+} B(\beta) = \infty,$$

imply that the necessary condition of the existence of the solution to (6) is vanishing of the derivative $\frac{\partial B}{\partial \beta}$ at some point. Assume first that $c > 0$. Then (7) has no zeros for $-1 < \beta < 0$. The derivative (7) vanishes at most once on the interval $0 < \beta < 1$. In fact, it vanishes exactly once if $0 < c < |w| - \tan^{-1}|w|$. Therefore, (6) has one solution in $0 < \beta < 1$.

Similarly, it can be shown that (6) has exactly one solution, when $\tan^{-1}|w| - |w| < c < 0$, and this solution β is between -1 and 0. If $c = 0$, (7) never vanishes in $\beta \in (-1, 1)$.

6.1.2. *The case* $\alpha_a = 0 \iff |\beta| = 1$. This happens when $r = |a_3|$. It is trivial that the equations

$$c = \tan^{-1} a_3 + a_3, \qquad |w| = |a_3|,$$

have a solution, which is unique, only if the endpoint satisfies the condition $c = \pm(|w| - \tan^{-1}|w|)$.

6.1.3. *The case* $\alpha_a > 0 \iff |\beta| > 1$*:* For this part, we need the following lemma.

Lemma 2. *For* $1 \le \beta \le \sqrt{1 + |w|^{-2}}$*, define*

$$\varphi_1 := \sin^{-1}\left(|w|\sqrt{\beta^2 - 1}\right), \qquad \varphi_2 := \sin^{-1}\left(\frac{|w|\beta}{\sqrt{1 + |w|^2}}\right).$$

Then $(\beta + \sqrt{\beta^2 - 1})(\beta\varphi_1(\beta) - \sqrt{\beta^2 - 1}\varphi_2(\beta))$ *monotonically increase from 0 to* $\frac{\pi}{2}$ *in the given interval.*

Proof. First, we observe that

$$\frac{\partial}{\partial\beta}\varphi_1(\beta) = \frac{\beta}{\sqrt{\beta^2 - 1}}\frac{|w|}{\sqrt{1 + |w|^2 - |w|^2\beta^2}}, \qquad \frac{\partial}{\partial\beta}\varphi_2(\beta) = \frac{|w|}{\sqrt{1 + |w|^2 - |w|^2\beta^2}}.$$

We come to the following chain

$$\frac{\partial}{\partial\beta}\left((\beta + \sqrt{\beta^2 - 1})(\beta\varphi_1(\beta) - \sqrt{\beta^2 - 1}\varphi_2(\beta))\right)$$

$$= \frac{\beta + \sqrt{\beta^2 - 1}}{\sqrt{\beta^2 - 1}\sqrt{1 + |w|^2 - |w|^2\beta^2}}\left(|w| - \sqrt{1 + |w|^2 - |w|^2\beta^2}(\beta + \sqrt{\beta^2 - 1})(\varphi_2 - \varphi_1)\right),$$

$$> \frac{\beta + \sqrt{\beta^2 - 1}}{\sqrt{\beta^2 - 1}\sqrt{1 + |w|^2 - |w|^2\beta^2}}\left(|w| - \tan^{-1}|w|\right) > 0,$$

In the above inequality, we used that $\sqrt{1 + |w|^2 - |w|^2\beta^2}(\beta + \sqrt{\beta^2 - 1})(\varphi_2(\beta) - \varphi_1(\beta))$ is decreasing, because

$$\frac{\partial}{\partial\beta}\left(\sqrt{1 + |w|^2 - |w|^2\beta^2}(\beta + \sqrt{\beta^2 - 1})(\varphi_2(\beta) - \varphi_1(\beta))\right)$$

$$= -\frac{1}{\sqrt{\beta^2 - 1}\sqrt{1 + |w|^2 - |w|^2\beta^2}}\left(|w| - \left(1 + |w|^2 - |w|^2\beta(\beta + \sqrt{\beta^2 - 1})\right)(\varphi_2(\beta) - \varphi_1(\beta))\right)$$

$$< -\frac{1}{\sqrt{\beta^2 - 1}\sqrt{1 + |w|^2 - |w|^2\beta^2}}\left(|w| - \tan^{-1}|w|\right) < 0.$$

□

Equations (2) are now written in a more complicated form

$$c = -\tan^{-1}\left(\frac{\beta}{\sqrt{\beta^2-1}}\tan\left(\frac{a_3\sqrt{\beta^2-1}}{\beta}\right)\right) - \pi\,\mathrm{sgn}\,(a_3)\left\lceil\frac{\sqrt{-\alpha_a}}{\pi}-\frac{1}{2}\right\rceil + a_3, \tag{8}$$

$$\left|\sin\left(\frac{a_3\sqrt{\beta^2-1}}{\beta}\right)\right| = |w|\sqrt{\beta^2-1}. \tag{9}$$

An immediate observation from (9), is that $|\beta| \leq \sqrt{1+|w|^{-2}}$. Let us introduce some notations. We write $\sqrt{-\alpha_a} = \pi k + \alpha', \alpha' \in (-\frac{\pi}{2}, \frac{\pi}{2}] \setminus \{0\}$, and also $\sigma_1 = \mathrm{sgn}\; a_3 = \mathrm{sgn}\;\beta$, $\sigma_2 = \mathrm{sgn}\;\alpha'$. If $k = 0$, then $\sigma_2 = 1$. Subsequently,

$$\sin(\sqrt{-\alpha_a}) = (-1)^k\sigma_2|w|\sqrt{\beta^2-1}, \qquad \cos(\sqrt{-\alpha_a}) = (-1)^k\sqrt{1+|w|^2-|w|^2\beta^2}.$$

So

$$a_3 = c + \sigma_2\tan^{-1}\left(\frac{|w|\beta}{\sqrt{1+|w|^2-|w|^2\beta^2}}\right) + \sigma_1\pi k = c + \sigma_2\varphi_2(\beta) + \sigma_1\pi k,$$

$$\alpha' = \sigma_2\tan^{-1}\left(\frac{|w|\sqrt{\beta^2-1}}{\sqrt{1+|w|^2-|w|^2\beta^2}}\right) = \sigma_2\varphi_1(\beta).$$

Then β is a solutions to the equation

$$k\pi + \sigma_2(|\beta| + \sqrt{\beta^2-1})(|\beta|\varphi_1 - \sqrt{\beta^2-1}|\varphi_2|) = |c|\sqrt{\beta^2-1}(|\beta| + \sqrt{\beta^2-1}), \tag{10}$$

where $\beta \in [-\sqrt{1+|w|^{-2}}, -1) \cup (1, \sqrt{1+|w|^{-2}}]$, and $\mathrm{sgn}\;\beta = \mathrm{sgn}\;c$. Also β can only have the value $\pm\sqrt{1+|w|^{-2}}$, when $\sigma_2 = 1$. From now we will just assume that $c > 0$, since the considerations for $c < 0$ are totally analogous. From Lemma 2, the left-hand side of (10) decreases or increases from $k\pi$ to $\frac{\pi}{2}(2k+\sigma_2)$. The derivative

$$\frac{\partial}{\partial\beta}\Big(((c+\varphi_2)\sqrt{\beta^2-1} - \beta\varphi_1)(\beta + \sqrt{\beta^2-1})\Big) \tag{11}$$

$$= \frac{(\beta+\sqrt{\beta^2-1})^2}{\sqrt{\beta^2-1}}\left(c - \frac{|w|(\beta-\sqrt{\beta^2-1})}{\sqrt{1+|w|^2-|w|^2\beta^2}} + (\varphi_2-\varphi_1)\right),$$

vanishes once if $|c| > |w| - \tan^{-1}|w|$, and otherwise it is negative. If $k = 0$, then (11) yields that a solution can occur only for $\frac{|w|-\tan^{-1}|w|}{\pi(\sqrt{|w|^2+1}-1)} \leq \frac{|c|}{\pi(\sqrt{|w|^2+1}-1)} \leq \frac{1}{2}$. Observe that

$$0 < \frac{|w| - \tan^{-1}|w|}{\sqrt{|w|^2+1}-1} < 1.$$

If $\sigma_2 = -1$ and $k > 0$, then there is a unique root if and only if, $k < \frac{|c|}{\pi(\sqrt{|w|^2+1}-1)} + \frac{1}{2}$. If $k > 0$ and $\sigma_2 = 1$, then there is clearly no solutions for $k \geq \frac{|c|(\sqrt{|w|^2+1}+1)}{\pi|w|^2}$, and from (11), it

follows that there is only one solution for $k < \frac{|c|}{\pi(\sqrt{|w|^2+1}-1)} - \frac{1}{2}$. Otherwise, let $\hat{\beta}$ be a root of the expression

$$c - \frac{|w|\left(\beta - \sqrt{\beta^2-1}\right)}{\sqrt{1+|w|^2-|w|^2\beta^2}} + (\varphi_2 - \varphi_1).$$

Then, there is a solution only if

$$k\pi + \left(\hat{\beta} + \sqrt{\hat{\beta}^2-1}\right)\left(\hat{\beta}\varphi_1 - \sqrt{\hat{\beta}^2-1}\varphi_2\right) \leq c\sqrt{\hat{\beta}^2-1}\left(\hat{\beta} + \sqrt{\hat{\beta}^2-1}\right).$$

If the equality is attained in the above inequality, then $\chi_k = \frac{|w|\sqrt{\hat{\beta}^2-1}}{\sqrt{1+|w|^2-|w|^2\hat{\beta}^2}}$, i.e., $\hat{\beta} = \frac{\sqrt{1+|w|^{-2}+\chi_k^{-2}}}{\sqrt{1+\chi_k^{-2}}}$. Hence, there is only one solution when

$$|c| = \sqrt{|w|^2+\chi_k^2+|w|^2\chi_k^2} - \chi_k \tan^{-1}\left(\frac{\sqrt{|w|^2+\chi_k^2+|w|^2\chi_k^2} - \chi_k}{1+\chi_k\sqrt{|w|^2+\chi_k^2+|w|^2\chi_k^2}}\right),$$

and there are two solutions when $|c|$ is between this value and $\frac{\pi}{2}(2k+1)(\sqrt{|w|^2+1}-1)$. All the cases has now been examined, so this ends the proof.

6.2. **Proof of Corollary 1.** Let us use the same notation as in the proof of Proposition 2. If $\tilde{g}(t)$ is a geodesic that goes from $\tilde{1}$ to (c, w) in a unit time interval, then $\ell(\tilde{g}) = r = \frac{a_3}{\beta}$. This implies (a)–(e) rather easily from the previous proof.

To prove (f), let g_{k,σ_2} be a geodesic corresponding to a choice of k and σ_2, with β_{k,σ_2}. The length becomes

$$\ell(\tilde{g}_{k,\sigma_2}) = \frac{|c| + \sigma_2|\varphi_2| + k\pi}{\beta_{k,\sigma_1}}.$$

Then, from the equation $\sqrt{\beta_{k,\sigma_2}^{-2}-1}(|c| + \sigma_2|\varphi_2| + k\pi) = \pi k + \sigma_2\varphi_1$, we know that

$$\ell(g_{k,\sigma_2})^2 = (|c| + k\pi + \sigma_2|\varphi_2(\beta_{k,\sigma_2})|)^2 + (\pi k + \sigma_2\varphi_1(\beta_{k,\sigma_2}))^2,$$

so

$$(|c| + k\pi)^2 + (\pi k)^2 < \ell(g_{k,1})^2 \leq \left(|c| + \left(k + \frac{1}{2}\right)\pi\right)^2 + \left(\pi\left(k + \frac{1}{2}\right)\right)^2,$$

$$\left(|c| + (k - \frac{1}{2})\pi\right)^2 + \left(\pi(k - \frac{1}{2})\right)^2 \leq \ell(g_{k,-1})^2 < (|c| + k\pi)^2 + (\pi k)^2.$$

It follows that $\tilde{g}_{1,-1}$ is minimal.

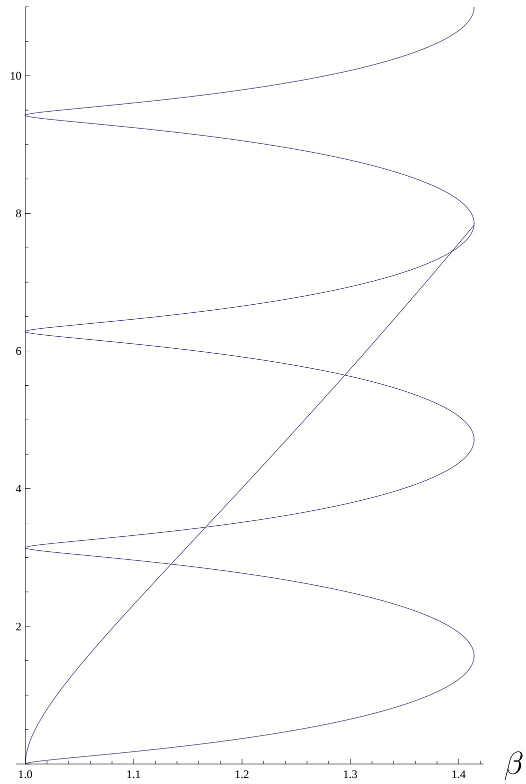

FIGURE 4. This figure essentially summarizes the results of the proof of Proposition 2. Here we present the case $|w| = 1$. The graphs with respect to β show the left-hand side of (10) where k varies from 0 to 3, and the right-hand side of (10), when $|c|$ is close to $\frac{5\pi}{2}(\sqrt{1+|w|^{-2}}-1)$. If $|c|$ is such that the graph of the right-hand side of (10) goes under all the above graphs, then the point corresponding to this choice of parameters is reached by a geodesic with $\alpha \geq 0$, see Corollary 1 (c). If $|c|$ is less than $\frac{\pi}{2}(\sqrt{|w|^{-2}+1}-1)$, but not underneath it, then this corresponds to the points in Corollary 1 (d). It is also easy to see that there can be no root or one root when $\sigma_2 = -1$, while if $\sigma_2 = 1$, the curve can either do not intersect, be tangent (when $|c|$ is in the cut locus), or intersect twice, like the curve in our figure.

6.3. **Proof of Proposition 5.** Since $a_3 > 0$ and $|a_2| < a_3$, let us denote $a_3 = r\cosh\theta$, and $a_2 = r\sinh\theta$. Then $\alpha_a = a_1^2 - r^2$. Similarly to the sub-Riemannian geodesics, we consider $\tilde{g}(0) = \tilde{1}, \tilde{g}(1) = (c, w)$, and look at how the final point defines the initial conditions. We are going to use the projection of (c, w) to $\mathrm{SU}(1,1)$. Along with the coordinates $z_1 = x_1 + iy_1$ and $z_2 = x_1 + iy_2$, we shall use new coordinates given by

$$u_1 = x_1 + x_2, \quad u_2 = x_1 - x_2, \quad v_1 = y_1 + y_2, \quad v_2 = y_1 - y_2.$$

Observe that $|z_1|^2 - |z_2|^2 = u_1u_2 + v_1v_2 = 1$. Then the projection of the geodesics to these coordinates becomes

$$\begin{aligned} u_1 &= e^{a_1}\left(C(\alpha_a) - a_1 S(\alpha_a)\right), & u_2 &= e^{-a_1}\left(C(\alpha_a) + a_1 S(\alpha_a)\right), \\ v_1 &= -rS(\alpha_a)e^{a_1-\theta}, & v_2 &= -rS(\alpha_a)e^{-(a_1-\theta)}. \end{aligned}$$

The relationship between the projected coordinates and the original ones on the coving space is

$$c = -\phi(a) + \tan^{-1}\left(\frac{2(u_2v_1 - u_1v_2)\sinh a_1}{e^{-a_1}(u_1^2+v_1^2+1) + e^{a_1}(u_2^2+v_2^2+1)}\right),$$

$$w = \frac{1}{2}(u_1 - u_2 + i(v_1 - v_2)).$$

Furthermore,

$$2C(a_1^2 - r^2) = u_1e^{-a_1} + u_2e^{a_1}, \tag{12}$$

$$2a_1S(a_1^2 - r^2) = -u_1e^{-a_1} + u_2e^{a_1}, \tag{13}$$

$$2r\sinh(\theta)S(a_1^2 - r^2) = v_1e^{-a_1} - v_2e^{a_1}, \tag{14}$$

$$2r\cosh(\theta)S(a_1^2 - r^2) = -v_1e^{-a_1} - v_2e^{a_1}. \tag{15}$$

First, let us consider the case when $a_1^2 - r^2 = -\pi^2k^2, k \in \mathbb{N}$. The only points that can be reached by this type of geodesics are $(-\pi k, x_2)$. Then

$$a_1 = (-1)^k \sinh^{-1} x_2, \qquad r = \sqrt{c^2 + (\sinh^{-1} x_2)^2},$$

and there are no restrictions on θ.

Otherwise, if $a_1^2 - r^2 \neq -\pi^2k^2$, then $v_1v_2 = 1 - u_1u_1 > 0$. It follow that

$$\theta = a_1 + \log\sqrt{\frac{v_2}{v_1}}, \quad \text{and} \quad r = \frac{2\sqrt{v_1v_2}|a_1|}{|u_2e^{a_1} - u_1e^{-a_1}|} \quad \text{if } a_1 \neq 0. \tag{16}$$

When $a_1 = 0$, we find r as $r = k + (-1)^{k+1}\sin^{-1}\left(\frac{(v_1+v_2)\sqrt{v_1v_2}}{|v_1|+|v_2|}\right)$, where $k = \lceil \frac{c}{\pi} - \frac{1}{2} \rceil$. So we only need to find the number of solutions to the equation (13) with respect to a_1, and then, the values of r and θ will be determined by (16).

Let us analyze a special case first. There is exactly one geodesic, satisfying $a_1 = 0$, when (c, w) is of the form

$$c = -\pi k + \operatorname{Arg}(\sqrt{1-s^2} - \operatorname{sgn}(s) i\sqrt{y_2^2 + s^2}), \quad w = iy_2,$$

where $y_2 \in \mathbb{R}, s \in (-1, 1] \setminus \{0\}$ or $s \in (0, 1]$ if $k = 0$. We continue the case by case analysis of the remaining possibilities. From (16), we know that

$$\alpha_a = \left(\frac{a_1}{\frac{u_2}{2}e^{a_1} - \frac{u_1}{2}e^{-a_1}}\right)^2\left(\left(\frac{u_2}{2}e^{a_1} + \frac{u_1}{2}e^{-a_1}\right)^2 - 1\right), \quad a_1 \neq 0.$$

6.3.1. $\alpha_a > 0$. A quick look at (14) and (15) yields that $v_1 < 0$ and $v_2 < 0$. Also from (12), we know that at least one of u_1 and u_2 is positive. We define $b = \sqrt{\left(\frac{u_2}{2}e^{a_1} + \frac{u_1}{2}e^{-a_1}\right)^2 - 1}$. Then,

$$e^{a_1} = \frac{\sqrt{b^2+1} + \sigma_1\sqrt{b^2+s^2}}{u_2} = \frac{u_1}{\sqrt{b^2+1} - \sigma_1\sqrt{b^2+s^2}}, \quad \sigma_1 = \pm 1.$$

Here, we denote $s = \sigma_1\sqrt{v_1 v_2}$. Notice, that the first part of the above equation is not valid if $u_2 = 0$, and the second part is not valid if $u_1 = 0$. The relation (13) implies sgn $a_1 = \sigma_1$.

If $\sigma_1 = 1$, then we know that sgn $u_2 = 1$. Furthermore, from (13), we obtain that b is a solution to the equation

$$(17) \quad \psi_{-1}(b) - u_2 = 0, \qquad \text{where } \psi_{-1}(b) := (\sqrt{b^2+1} + \sqrt{b^2+s^2})e^{-\frac{\sqrt{b^2+s^2}}{b}\log(b+\sqrt{b^2+1})},$$

and ψ_{-1} increases from $(1+s)e^{-s}$ to 1. So, a solution exists only if $(1+s)e^{-s} < u_2 < 1$. Then the points, which are connected by these types of geodesics are

$$c = \operatorname{Arg}\left(\frac{u_2^2 + 1 - s^2}{2u_2} - i\sqrt{y_2^2 + s^2}\right), \qquad w = \frac{1 - s^2 - u_2^2}{2u_2} + iy_2,$$

where $(1+s)e^{-s} < u_2 < 1$. To get some idea of the meaning the above equations, notice that sgn $(1-s) =$ sgn u_1, so $x_1 = \frac{1}{2}(\frac{1-s^2}{u_2} + u_2)$ is an increasing function of u_2 when $s > 1$. Also, $x_2 = \frac{1}{2}(\frac{1-s^2}{u_2} - u_2)$ is a decreasing function of u_2 when $s < 1$. Using this information, we know that there are only geodesic of these types connecting the points of the form

$$c = \operatorname{Arg}\left(x_1 - i\sqrt{s^2 + y_2^2}\right), \qquad w = x_2 + iy_2,$$

with the identity of the group, where $y_2 \in \mathbb{R}, s > 0$ and x_1, x_2 satisfy one of the following conditions

- $x_2 \in (-\frac{s^2}{2}, \sinh s - s\cosh s)$, $x_1 = \sqrt{x_2^2 + 1 - s^2}$, if $0 < s < 1$;
- $x_1 = -x_2 \in (e^{-1}, \frac{1}{2})$, if $s = 1$;
- $x_1 \in (\cosh s - s\sinh s, 1 - \frac{s^2}{2}), x_2 = -\sqrt{x_1^2 + s^2 - 1}$, if $s > 1$.

We can rewrite the above cases as the condition $x_2 = -\sqrt{x_1^2 + 1 - s^2}$, $x_1 \in (\cosh s - s\sinh s, 1 - \frac{s^2}{2})$.

When $\sigma_1 = -1$, we know that sgn $u_1 = 1$, and b is a solution to the equation $\psi_{-1}(b) = \log u_1$. So the points reachable by these types of geodesics are

$$c = \operatorname{Arg}\left(x_1 - i\sqrt{y_2^2 + s^2}\right), \qquad w = x_2 + iy_2,$$

where $s < 0, y_2 \in \mathbb{R}$, and x_1, x_2 satisfy one of the following conditions

- $x_2 \in (\sinh s - s\cosh s, \frac{s^2}{2}), x_1 = \sqrt{x_2^2 + 1 - s^2}$, if $0 > s > -1$;
- $x_1 = x_2 \in (e^{-1}, \frac{1}{2})$ if $s = -1$;
- $x_1 \in (\cosh s - s\sinh s, 1 - \frac{s^2}{2}), x_2 = \sqrt{x_1^2 + s^2 - 1}$, if $s < -1$.

Notice that the only that changes with the sign of σ_1, is the sign of x_2.

6.3.2. $\alpha_a = 0$. From (12) and (13), it follows that there are geodesics only of this type connecting the initial point with every point of the form

$$\left\{ \left(\mathrm{Arg} \left(\cosh s - s \sinh s - i\sqrt{s^2 + y_2^2} \right), \sinh s - s\cosh s + iy_2 \right) \middle| s, y_2 \in \mathbb{R}, s \neq 0 \right\}.$$

6.3.3. $\alpha_a < 0$. We introduce the notation $\sqrt{-\alpha_a} = \pi k + \alpha', \alpha' \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right] \setminus \{0\}$. Let $\sigma_2 = \mathrm{sgn}\ \alpha'$. Notice that $\mathrm{sgn}\ v_1 = \mathrm{sgn}\ v_2 = (-1)^{k+1}\sigma_2$, and hence, $\mathrm{sgn}\ y_1 = (-1)^{k+1}\sigma_2$. Now define

$$\varphi_1 := \tan^{-1}\left(\frac{r\cosh\theta}{\sqrt{-\alpha}}\tan\alpha'\right), \qquad \varphi_2 := \tan^{-1}\left(\frac{2(u_2v_1 - u_1v_2)\sinh a_1}{e^{-a_1}(u_1^2 + v_1^2 + 1) + e^{a_1}(u_2^2 + v_2^2 + 1)}\right),$$

so that $c = -\pi k - \varphi_1 + \varphi_2$. From the fact that $\sin c = (-1)^{k+1}\sin(\varphi_1 - \varphi_2) = y_1$, we know that $\mathrm{sgn}\ (\varphi_1 - \varphi_2) = \mathrm{sgn}\ \varphi_1 = \sigma_2$, which yields that $k = -\lceil \frac{c}{\pi} \rceil - \frac{1-\sigma_2}{2}$. Now write $c = 2n\pi + \mathrm{Arg}(z_1)$. Then, $2n = k + \frac{(1+(-1)^{k+1})\sigma_2}{2}$, so

$$c = \begin{cases} -\pi k + \mathrm{Arg}(-z_1), & \text{if } k \text{ is odd;} \\ -k\pi + \mathrm{Arg}(z_1), & \text{if } k \text{ is even.} \end{cases}$$

Given a value of k, the value of c is completely defined by z_1. We also know that if two geodesics has the same endpoint, then the maximal difference in their value of k is 1.

We need to find a_1 as a solution to

$$\frac{2a_1\sqrt{1 - \left(\frac{u_2}{2}e^{a_1} + \frac{u_1}{2}e^{-a_1}\right)^2}}{u_2e^{a_1} - u_1e^{-a_1}} = \sigma_2(-1)^k\pi k + (-1)^k\sin^{-1}\left(\sqrt{1 - \left(\frac{u_2}{2}e^{a_1} + \frac{u_1}{2}e^{-a_1}\right)^2}\right).$$

Let us denote $b = \sqrt{1 - \left(\frac{u_2}{2}e^{a_1} + \frac{u_1}{2}e^{-a_1}\right)^2}$. It follows that

$$e^{a_1} = \frac{(-1)^k(\sqrt{1-b^2} + \sigma_1\sqrt{v_1v_2 - b^2})}{u_2} = \frac{u_1}{(-1)^k(\sqrt{1-b^2} - \sigma_1\sqrt{v_1v_2 - b^2})}, \quad \sigma_1 = \pm 1.$$

As in the previous cases, there are some obvious restrictions on the values of $u_2 = 0$ or $u_1 = 0$. Here we used that $\mathrm{sgn}\ \left(\frac{u_2}{2}e^{a_1} + \frac{u_1}{2}e^{-a_1}\right) = (-1)^k$, because of (12). From (13), we know that $\sigma_1 = \sigma_2\,\mathrm{sgn}\ a_1$.

If $\sigma_1 = 1$, then $\mathrm{sgn}\ u_2 = (-1)^k$, and so b is a solution to

$$\text{(18)} \qquad \psi_k(b) = (-1)^k u_2, \quad \psi_k(b) := \left(\sqrt{1-b^2} + \sqrt{s^2 - b^2}\right)e^{-\frac{\sqrt{s^2-b^2}}{b}(\sigma_2\pi k + \sin^{-1} b)},$$

where $0 < b \leq \min\{1, |s|\}$. Again we use the notation $s = \sigma_1\sqrt{v_1v_2}$. The upper bound for b is strict, unless $|q_1| > 1$ and $\sigma_2 = 1$, because α' can be $\frac{\pi}{2}$, but not $-\frac{\pi}{2}$. Moreover, from (13), $b = |s|$ implies that either $a_1 = 0$ or $\sin\sqrt{-\alpha_a} = 0$, the case which already excluded. Similarly, when $\sigma_1 = -1$, so b is a solution to $\psi_k(b) = (-1)^k u_1$.

We investigate when a solution can exist.

- We start with the special case $k = 0$. Then ψ_0 decreases from $(1 + |s|)e^{-|s|}$ to $\sqrt{s^2-1}e^{-\frac{\pi}{2}\sqrt{s^2-1}}$ for $s \geq 1$, and from $(1 + |s|)e^{-|s|}$ to $\sqrt{1-s^2}$ for $|s| < 1$.

- If $\sigma_2 = -1$, then ψ_k decreases from ∞ to $\sqrt{1-s^2}$ for $|s| < 1$, or to
$$\sqrt{s^2-1}\exp\left(\frac{(2k-1)\pi}{2}\sqrt{s^2-1}\right)$$
otherwise.
- If $\sigma_2 = 1$, then ψ_k increases from 0 to $\sqrt{1-s^2}$ for $|s| \leq \omega_k$. Continuing with $\sigma_2 = 1$, if $|s| > \omega_k$, then ψ_k increases from 0 to $(\sqrt{1-\omega_k^2}+\sqrt{s^2-\omega_k^2})\exp\left(-\sqrt{\frac{s^2-\omega_k^2}{1-\omega_k^2}}\right)$, and decreases to $\sqrt{1-s^2}$ for $|s| < 1$, or to $\sqrt{s^2-1}\exp\left(-\frac{(2k+1)\pi}{2}\sqrt{s^2-1}\right)$ otherwise.

Further, if $\sigma_1 = 1$, then sgn $u_2 = (-1)^k$, and sgn $u_1 = (-1)^k$sgn $(1-s^2)$, and the same holds when $\sigma_1 = -1$, interchanging u_1 and u_2. Also, from
$$x_1 = \frac{1}{2}\left(u_1 + \frac{1-s^2}{u_1}\right) = \frac{1}{2}\left(\frac{1-s^2}{u_2} + u_2\right), \quad x_2 = \frac{1}{2}\left(u_1 - \frac{1-s^2}{u_1}\right) = \frac{1}{2}\left(\frac{1-s^2}{u_2} - u_2\right),$$
we know that x_1 is an increasing function of u_1 and u_2, when $|s| > 1$. When $|s| < 1$, x_2 increases in u_1 and decreases in u_2. As a result, we know which point is reachable by which type of geodesics. We only show the results for $s > 0$, since for $s < 0$, the results are the same, only with different sign for x_2.

- If $k = 0$, then
  - if $0 < s < 1$, then $x_1 = \sqrt{x_2^2 + 1 - s^2}$, $x_2 \in (\sinh s - s\cosh s, 0)$;
  - if $1 \leq s$, $x_2 = -\sqrt{x_1^2 + s^2 - 1}$, $x_1 \in [f_+(s;0), \cosh s - s\sinh s)$.
- If $k > 0$ and $\sigma_2 = -1$, then
  - if $0 < s < 1$, $(-1)^k x_1 = \sqrt{x_2^2 + 1 - s^2}$, and $(-1)^k x_2 \in (-\infty, 0)$;
  - if $s \geq 1$, $(-1)^k x_2 = -\sqrt{x_1^2 + s^2 - 1}$, and $(-1)^k x_1 \in (f_-(s;k), \infty)$.
- If $k > 0$ and $\sigma_2 = 1$, then there is one geodesic (list 1)
  - if $0 < s \leq \omega_k$, $(-1)^k x_1 = \sqrt{x_2^2 + 1 - s^2}$, and $(-1)^k x_2 \in (0, \infty)$;
  - if $\omega_k < s < 1$, $(-1)^k x_1 = \sqrt{x_2^2 + 1 - s^2}$, and $(-1)^k x_2 \in [0, \infty)$ $\cup \left\{\sqrt{1-\omega_k^2}\sinh\left(\sqrt{\frac{s^2-\omega_k^2}{1-\omega_k^2}}\right) - \sqrt{s^2-\omega_k^2}\cosh\left(\sqrt{\frac{s^2-\omega_k^2}{1-\omega_k^2}}\right)\right\}$;
  - if $s = 1$ and $(-1)^k x_1 = -(-1)^k x_2 = \sqrt{1-\omega_k^2}e^{-1}$;
  - if $s > 1$, $(-1)^k x_2 = -\sqrt{x_1^2 + s^2 - 1}$, and
$$\begin{aligned}(-1)^k x_1 \quad &\in \quad \left(-\infty, -\sqrt{s^2-1}\sinh\left(\frac{(2k+1)\pi}{2}\sqrt{s^2-1}\right)\right) \\ &\cup \quad \left\{\sqrt{1-\omega_k^2}\cosh\left(\sqrt{\frac{s^2-\omega_k^2}{1-\omega_k^2}}\right) - \sqrt{s^2-\omega_k^2}\sinh\left(\sqrt{\frac{s^2-\omega_k^2}{1-\omega_k^2}}\right)\right\},\end{aligned}$$

  and two geodesics (list 2)
  - if $\omega_k < s < 1$, $(-1)^k x_1 = \sqrt{x_2^2 + 1 - s^2}$, and
$$(-1)^k x_2 \in \left(\sqrt{1-\omega_k^2}\sinh\left(\sqrt{\frac{s^2-\omega_k^2}{1-\omega_k^2}}\right) - \sqrt{s^2-\omega_k^2}\cosh\left(\sqrt{\frac{s^2-\omega_k^2}{1-\omega_k^2}}\right), 0\right);$$

– if $s = 1$, and $(-1)^k x_1 = -(-1)^k x_2 \in (0, \sqrt{1-\omega_k^2}e^{-1})$;
– if $s > 1$, $(-1)^k x_2 = -\sqrt{x_1^2 + s^2 - 1}$, and

$$(-1)^k x_1 \in \quad \Bigg[[-\sqrt{s^2-1}\sinh\left(\frac{(2k+1)\pi}{2}\sqrt{s^2-1}\right),$$
$$\sqrt{1-\omega_k^2}\cosh\left(\sqrt{\frac{s^2-\omega_k^2}{1-\omega_k^2}}\right) - \sqrt{s^2-\omega_k^2}\sinh\left(\sqrt{\frac{s^2-\omega_k^2}{1-\omega_k^2}}\right)\Bigg).$$

The above list can be rewritten so the interval is determined for x_1 only. Note that the term "one geodesic" and "two geodesics" is a bit misleading, since, if we include both signs of s, list 1 and list 2 are not disjoint.

6.3.4. *Summary.* We start by defining some sets to sum up the information obtained so far:

- $\tilde{A}_0$ consist of all points
$$c = \mathrm{Arg}\left(x_1 - i\sqrt{s^2+y_2^2}\right), \quad w = \pm\sqrt{x_1^2+s^2-1} + iy_2,$$
where $s, y_2 \in \mathbb{R}$, $s > 0$, $x_1 \in (f_+(s;0), 1-\frac{s^2}{2})$, when $0 < s \leq 1$, and $x_1 \in [f_+(s;0), 1-\frac{s^2}{2})$, when $s > 1$. They are the points that are reachable by geodesics with $\alpha_a \geq 0$, and with $\alpha_a < 0$, $k = 0$.
- $A_k^0, k \in \mathbb{N}_0$, consist of the points
$$c = -\pi k + \mathrm{Arg}(\sqrt{1-s^2} - i\sqrt{y_2^2+s^2}), \quad w = iy_2,$$
where $y_2 \in \mathbb{R}, s \in (0,1]$, and define $A_{-k}^0, k \in \mathbb{N}$, by.
$$c = -\pi k + \mathrm{Arg}(\sqrt{1-s^2} + i\sqrt{y_2^2+s^2}), \quad w = iy_2,$$
where $y_2 \in \mathbb{R}, s \in (0,1)$. They are the points reachable by geodesics with $a_1 = 0$.
- $A_k^-, k \in \mathbb{N}$, consist of all points
$$c = -(k-1)\pi + \mathrm{Arg}\left(x_1 - i\sqrt{s^2+y_2^2}\right), \quad w = \pm\sqrt{x_1^2+s^2-1} + iy_2,$$
where $s, y_2 \in \mathbb{R}$, $s > 0$ and $x_1 \in (-\infty, -f_-(s;k))$. They are the points reachable by geodesic with a certain choice of k, and with $\sigma_2 = -1$.
- $A_k^+, k \in \mathbb{N}$, consist of all points
$$c = -k\pi + \mathrm{Arg}\left(x_1 - i\sqrt{s^2+y_2^2}\right), \quad w = \pm\sqrt{x_1^2+s^2-1} + iy_2,$$
where $s, y_2 \in \mathbb{R}$, $s > 0, s \neq 1$, and
  – $x_1 \in (-\infty, f_+(s;k)) \cup \{F(s,\omega_k)\}$, when $s > 1$;
  – $x_1 = F(1,\omega_k)$ if $s = 1$;
  – $x_1 \in [f_+(s;k), \infty) \setminus \{F(s,\omega_k\}$, when $\omega_k < s < 1$;
  – $x_1 \in (f_+(s;k), \infty)$, when $0 < s \leq \omega_k$.

They are the points from list 1, for $\sigma_2 = 1$ and for a choice of k (except for the points $x_1 = F(s, \omega_k)$, for $\omega_k < s < 1$, that appear twice and which are therefore included in $\tilde{B}_k$).

- $\tilde{B}_k, k \in \mathbb{N}$, consist of all points

$$c = -k\pi + \operatorname{Arg}\left(x_1 - i\sqrt{s^2 + y_2^2}\right), \quad w = \pm\sqrt{x_1^2 + s^2 - 1} + iy_2,$$

where $s, y_2 \in \mathbb{R}$, $s > \omega_k$ and
  - $x_1 \in (f_+(s;k), F(s, \omega_k)]$ when $\omega_k < s < 1$;
  - $x_1 \in (0, F(1, \omega_k))$ when $s = 1$;
  - $x_1 \in [f_+(s;k), F(s, \omega_k))$ when $s > 1$.

They are the points from list 2 for a certain choice of k and with $\sigma_2 = 1$.

Notice that A_0^0 is on the boundary of $\tilde{A}_0$, A_{-k}^0 is on the boundary of A_k^- and A_k^0 is on the boundary of A_k^+ for $k > 0$. The result follows merely by comparing the sets and counting number of the sets in which a value of x_1 appears. Points in $\tilde{B}_k$ are counted twice. We use $\tilde{A}_0$, A_0^0, A_{-1}^0, A_1^-, and A_{-k}^0 to define Ω_0, and use A_k^+, A_k^0, A_{k+1}^-, A_{-k-1}^0, and $\tilde{B}_k$ to define A_k, B_k, and C_k.

This ends the proof.

References

[1] A. Agrachev, *Exponential mapping for contact sub-Riemannian structures*, Journal of Dynamical and Control Systems, **2**, (1996), no. 3, 321–358.

[2] A. Agrachev, Yu. Sachkov, *Control theory from the geometric viewpoint*, Encyclopaedia of Math. Sci., 87. Control Theory and Optimization, II. Springer-Verlag, Berlin, 2004, 412 pp.

[3] A. Bellaïche, J.-J. Risler, *Sub-Riemannian Geometry*, Birkhäuser Verlag, Basel, 1996, 393 pp.

[4] U. Boscain, F. Rossi, *Invariant Carnot-Caratheodory metric on S^3, $SO(3)$, $SL(2)$ and lens spaces*, Siam J. Control Optim. **47** (2008), no. 4, 1851–1878.

[5] O. Calin, D.-C. Chang, *Sub-Riemannian geometry. General theory and examples*, Cambridge Univ. Press, 2009, 384 pp.

[6] O. Calin, D.-C. Chang, I. Markina, *Sub-Riemannian geometry of the sphere S^3*, Canadian J. Math. **61** (2009), no. 4, 821–839.

[7] S. Carlip, *Conformal field theory, $(2+1)$-dimensional gravity and the BTZ black hole*, Classical Quantum Gravity **22** (2005), no. 12, R85–R123.

[8] D.-C. Chang, I. Markina, A. Vasil'ev, *Sub-Lorentzian geometry on anti-de Sitter space*, J. Math. Pures Appl. **90** (2008), no.1, 82–110.

[9] D.-C. Chang, I. Markina, A. Vasil'ev, *Sub-Riemannian geodesics on the 3-D sphere*, Compl. Analysis Operator Theory, **3** (2009), no. 2, 361–377.

[10] W. L. Chow. *Über Systeme von linearen partiellen Differentialgleichungen erster Ordnung*, Math. Ann., **117** (1939), 98–105.

[11] M. Grochowski, *Geodesics in the sub-Lorentzian geometry*, Bull. Polish Acad. Sci. Math. **50** (2002), no. 2, 161–178.

[12] M. Grochowski, *On the Heisenberg sub-Lorentzian metric on* $\mathbb{R}^3$, Geometric Singularity Theory, Banach Center Publ., Polish Acad. Sci. **65** (2004), 57–65.

[13] M. Grochowski, *Reachable sets for the Heisenberg sub-Lorentzian structure $\mathbb{R}^3$. An estimate for the distance function*, J. Dynamical and Control Systems, **12** (2006), no. 2, 145–160.

[14] A. Hurtado, C. Rosales, *Area-stationary surfaces inside the sub-Riemannian three-sphere*, Math. Ann. **340** (2008), 675–708.

[15] V. Jurdjevic, *Geometric control theory*, Cambridge Studies in Adv. Math., 52. Cambridge Univ. Press, Cambridge, 1997, 492 pp.

[16] V. Jurdjevic, *Optimal control, geometry, and mechanics*, Mathematical Control Theory, Springer, New York, 1999, 227–267.

[17] A. Korolko, I. Markina, *Nonholonomic Lorentzian geometry on some* $\mathbb{H}$*-type groups*, J. Geom. Anal. **19** (2009), 864–889.

[18] W. Liu, H. J. Sussman, *Shortest paths for sub-Riemannian metrics on rank-two distributions*, Mem. Amer. Math. Soc. **118** (1995), no. 564, 104 pp.

[19] R. Montgomery, *A tour of subriemannian geometries, their geodesics and applications*, Mathematical Surveys and Monographs, **91**. American Mathematical Society, Providence, RI, 2002, 259 pp.

[20] P. K. Rashevskiĭ, *About connecting two points of complete nonholonomic space by admissible curve*, Uch. Zapiski Ped. Inst. K. Liebknecht **2** (1938), 83–94.

[21] R. S. Strichartz, *Sub-Riemannian geometry*, J. Differential Geom. **24** (1986), no. 2, 221–263.

[22] R. S. Strichartz, *Corrections to: "Sub-Riemannian geometry" J. Differential Geom.* **24** *(1986), no. 2, 221–263;* J. Differential Geom. bf 30 (1989), no. 2, 595–596.

[23] A. M. Vershik, V. Ya. Gershkovich, *Geodesic flows on* $\mathrm{SL}(2,\mathbb{R})$ *with nonholonomic restrictions*, Zap. Nauchn. Semin. LOMI, **155** (1986), 7–17.

[24] E. Witten, *String theory and black holes*, Phys. Rev. D **44** (1991), no. 2, 314–324.

E. Grong and A. Vasil'ev:
Department of Mathematics
University of Bergen
P.O. Box 7803
Bergen N-5020
Norway

E-mail address: `erlend.grong@math.uib.no`
E-mail address: `alexander.vasiliev@math.uib.no`